БЕЛКООПСОЮЗ
УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ
«БЕЛОРУССКИЙ ТОРГОВО-ЭКОНОМИЧЕСКИЙ УНИВЕРСИТЕТ ПОТРЕБИТЕЛЬСКОЙ КООПЕРАЦИИ»

Кафедра высшей математики

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Пособие
по подготовке к тестированию
для студентов заочной формы обучения
экономических специальностей

Гомель 2009

ББК 22.171
УДК 519.2
Т 33

Авторы-составители: Л. П. Авдашкова, канд. физ.-мат. наук, доцент;
Т. Ф. Калмыкова, канд. техн. наук, доцент;
И. А. Кузменкова, канд. физ.-мат. наук, доцент;
Н. Г. Лопухова, канд. физ.-мат. наук, доцент;
С. А. Мокеева, канд. физ.-мат. наук, доцент;
О. А. Мокеева, ассистент

Рецензенты: А. А. Бабич, канд. физ.-мат. наук, доцент, заведующий кафедрой высшей математики Гомельского государственного технического университета имени П. О. Сухого;
В. И. Тютин, канд. физ.-мат. наук, доцент, заведующий кафедрой высшей математики Белорусского торгово-экономического университета потребительской кооперации

Рекомендовано к изданию научно-методическим советом учреждения образования «Белорусский торгово-экономический университет потребительской кооперации». Протокол № 5 от 14 апреля 2009 г.

Т 33 **Теория** вероятностей и математическая статистика : пособие по подготовке к тестированию для студентов заочной формы обучения экономических специальностей / авт.-сост.: Л. П. Авдашкова [и др.]. – Гомель : учреждение образования «Белорусский торгово-экономический университет потребительской кооперации», 2009. – 136 с.
ISBN 978-985-461-716-9

ББК 22.171
УДК 519.2

ISBN 978-985-461-716-9

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Знание математики необходимо студентам экономических вузов при изучении фундаментальных и специальных экономических дисциплин. Теория вероятностей и математическая статистика позволяют освоить методы изучения, обобщения и прогнозирования экономической информации.

Данное пособие составлено в соответствии с программой дисциплины «Теория вероятностей и математическая статистика» в экономическом вузе. Оно состоит из девяти разделов, охватывающих основные темы курса: «Случайные события», «Случайные величины», «Элементы математической статистики». В начале каждого раздела приведены необходимые теоретические сведения, даны решения типовых задач, приведены примеры тестовых заданий, в которых необходимо закончить фразу, выбрав правильный вариант ответа.

Цель данного пособия – развитие и закрепление навыков решения тестовых заданий, освоения студентами необходимого математического аппарата, овладения приемов исследования и решения математических формализованных задач.

ПРОГРАММА ДИСЦИПЛИНЫ

Тема 1. Случайные события и вероятность

Случайные события и операции над ними. Алгебра событий. Частота и вероятность. Классическое определение вероятности. Геометрические вероятности и статистическая вероятность. Теоремы сложения и умножения вероятностей. Условная вероятность. Независимость событий. Формулы полной вероятности и Байеса.

Последовательность независимых повторных испытаний. Формула Бернулли. Наивероятнейшее число успехов в схеме Бернулли. Предельные теоремы в схеме Бернулли.

Тема 2. Случайные величины и законы их распределения

Случайные величины и их классификация. Дискретные и непрерывные величины. Законы распределения случайных величин. Функция распределения случайных величин и ее свойства. Плотность распределения непрерывной случайной величины и ее свойства. Математическое ожидание и дисперсия случайной величины. Мода и медиана. Моменты случайной величины. Асимметрия и эксцесс. Функции случайных величин.

Биномиальный закон распределения. Закон Пуассона. Геометрическое и гипергеометрическое распределения. Равномерное распределение. Показательное распределение. Нормальный закон распределения. Функция Лапласа. Распределения «хи-квадрат», Стьюдента и Фишера.

Многомерные случайные величины. Зависимые и независимые случайные величины. Корреляционный момент и коэффициент корреляции.

Тема 3. Закон больших чисел

Неравенства Маркова и Чебышева. Теоремы Чебышева и Бернулли. Центральная предельная теорема.

Тема 4. Основы математической статистики

Предмет математической статистики. Генеральная и выборочная совокупности. Вариационный ряд и его характеристики. Точечное и интервальное оценивание параметров генеральной совокупности. Предельная ошибка и необходимый объем выборки.

Статистические гипотезы. Уровень значимости и мощность критерия. Проверка статистических гипотез. Критерии согласия Пирсона и Колмогорова.

Основные понятия дисперсионного анализа. Однофакторный и двухфакторный дисперсионный анализ.

Модели и основные понятия корреляционного и регрессионного анализа. Линейная корреляционная зависимость и линии регрессии. Проверка значимости уравнения и коэффициентов уравнения регрессии. Ранговая корреляция.

1. ОСНОВНЫЕ ПОНЯТИЯ И ТЕОРЕМЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

1.1. Классификация событий. Действия над событиями

Предметом теории вероятностей является изучение вероятностных закономерностей массовых однородных случайных событий.

Знание закономерностей, которым подчиняются массовые случайные события, позволяет предвидеть, как эти события будут протекать.

Пример 1.1. Монету бросают в одних и тех же условиях. Нельзя наперед определить результат одного бросания монеты, но можно предсказать, причем с небольшой погрешностью, число появлений «герба», если монета будет брошена достаточно большое число раз.

Испытанием называется осуществление определенной совокупности условий.

Пример 1.2. Испытанием является стрельба по мишени, бросание монеты, извлечение шара из ящика, бросание игрального кубика (шестигранного кубика, на гранях которого различное число очков от 1 до 6).

Событие – результат испытания. События, как правило, обозначаются заглавными латинскими буквами A, B, C ….

Пример 1.3. При стрельбе по мишени событиями будут «попадание» или «промах», при бросании монеты – «герб» или «надпись» на верхней ее стороне; появление бракованного изделия при выборке из множества готовых изделий и т. д.

Исходы испытания – события, каждое из которых может произойти в результате испытания.

Пример 1.4. В ящике находятся красные, зеленые шары. Наудачу извлекают один шар из ящика. Испытанием является извлечение шара из ящика. Исходы – появление красного шара, появление зеленого шара.

Тест 1.1. В ящике находятся белые, синие, желтые шары. Наудачу извлекают один шар из ящика. Испытанием является:
1) появление цветного шара;
2) извлечение шара из ящика;
3) появление белого шара;
4) появление синего шара;
5) появление желтого шара.

События, происходящие в окружающем нас мире, можно разделить на три вида: достоверные, невозможные и случайные.

Достоверное событие – это такое событие, которое всегда произойдет в результате испытания. Оно обозначается через E.

Невозможное событие – это такое событие, которое не может произойти в результате испытания. Оно обозначается через U.

Пример 1.5. В партии все изделия стандартные. Извлечение из нее стандартного изделия – событие достоверное, а извлечение при тех же условиях бракованного изделия – событие невозможное.

Случайным называется событие, которое может произойти или не произойти в результате испытания.

Пример 1.6. В ящике находятся 5 голубых и 3 красных шара, одинаковых по размеру и весу. Событие «из ящика извлечен голубой шар» является случайным, так как оно может произойти, а может и не произойти, поскольку в ящике имеются не только голубые, но и красные шары.

Виды случайных событий: несовместные, совместные, равновозможные, единственно возможные, противоположные, элементарные, составные.

Два события называются *несовместными*, если появление одного из них исключает возможность появления другого события в одном и том же испытании, в противном случае события называются *совместными*.

Пример 1.7. Бросается монета. Появление «надписи» исключает появление «герба», и наоборот. Поэтому события «появилась надпись» и «появился герб» – несовместные события.

Пример 1.8. В аудиторию вошел человек. События «в аудиторию вошел человек старше 35 лет» и «в аудиторию вошел мужчина» – совместные, поскольку в аудиторию может войти мужчина старше 35 лет.

События называются *равновозможными*, если каждое из них не является более возможным, чем другое.

Пример 1.9. Испытание – подбрасывание монеты. События появление «герба» или «надписи» равновозможные. Так как, если монета «правильная», выполнена симметрично, то нет никаких оснований считать «появление герба» при подбрасывании монеты событием объективно более возможным, чем «появление надписи».

События называются *единственно возможными*, если в результате испытания обязательно должно произойти, хотя бы одно из них.

Пример 1.10. В ящике содержатся белые, черные и красные шары. Извлекаем из ящика шар, который может оказаться белым (событие A), черным (событие B) или красным (событие C). По определению эти три события A, B, C – единственно возможные.

Противоположным событию A называется событие $\overline{A}$, которое происходит тогда и только тогда, когда не происходит событие A (т. е. $\overline{A}$ означает, что событие A не наступило).

Пример 1.11. Производится выстрел по мишени. События «попадание» и «промах» являются противоположными. Если A – попадание, то $\overline{A}$ – промах. Если же $\overline{A}$ – промах, то A – попадание.

Элементарным событием (элементарным исходом) называется каждое событие, которое может наступить в результате испытания.

Составным событием (составным исходом) называется событие, которое можно представить в виде совокупности элементарных исходов (событий) этого испытания.

Пример 1.12. Подбрасывается игральный кубик. Испытанием является подбрасывание кубика. Элементарное событие – выпадение грани с очком 1. Составное событие – выпадение грани с нечетным числом очков.

Тест 1.2. В ящике находятся голубые шары. Событие «из ящика извлечен голубой шар» является:
1) противоположным;
2) совместным;
3) достоверным;
4) невозможным;
5) равновозможным.

Тест 1.3. Произведено два выстрела по мишени. События: A_1 – «два попадания», A_2 – «только одно попадание», A_3 – «ни одного попадания» являются:
1) совместными;
2) равновозможными;
3) несовместными;
4) противоположными.

Тест 1.4. При одном бросании игрального кубика элементарным является событие:
1) «появление четырех очков»;
2) «появление нечетного числа очков»;
3) «появление числа очков, кратного 3»;
4) «появление числа очков, меньшего пяти»;
5) «появление 0 очков».

Несколько событий образуют *полную группу*, если они являются единственно возможными и несовместными исходами испытания.

Пример 1.13. Бросается игральный кубик. События, заключающиеся в том, что на верхней грани кубика появится 1, 2, 3, 4, 5, 6 очков, образуют полную группу событий, так как в результате испытания кубик обязательно упадет какой-нибудь гранью вверх, а значит, произойдет одно из указанных событий.

Суммой $A + B$ двух событий A и B называется событие, состоящее в появлении хотя бы одного из них.

Произведением двух событий A и B называют событие AB, состоящее в совместном появлении этих событий.

Пример 1.14. В ящике находятся белые, синие и черные шары. Извлекается один шар. Возможны следующие события: A – «извлечен белый шар», B – «извлечен синий шар», C – «извлечен черный шар». Событие $A + B$ означает, что произошло событие «извлечен не черный шар», а событие $B + C$ – «извлечен не белый шар».

Пример 1.15. Бросается игральный кубик. Рассмотрим следующие события: A – «число выпавших очков меньше 5», B – «число выпавших очков больше 2», C – «число выпавших очков четное». Тогда событие ABC заключается в том, что выпало 3 очка.

Тест 1.5. Из корзины, содержащей красные, желтые и белые розы, выбирается один цветок. Пусть события: A – «выбрана красная роза», B – «выбрана желтая роза». Тогда событие C – «выбрана белая роза» равно:

1) $\overline{A}$;
2) $A + B$;
3) $\overline{A} + \overline{B}$;
4) $\overline{A + B}$;
5) AB.

Тест 1.6. Пусть A, B, C – три произвольных события. Событие «произошло, по крайней мере, одно из событий A, B, C» равно:

1) $A + B + C$;
2) $\overline{A}\overline{B}C$;
3) ABC;
4) $AB + AC + BC$;
5) $\overline{A}BC + A\overline{B}C + AB\overline{C}$.

1.2. Понятие вероятности

В повседневной жизни в разговоре часто используется слово «вероятный». Например, «к вечеру, вероятно, пойдет дождь», «вероятнее всего он опоздает». При употреблении этого слова интуитивно оценивается возможность наступления того или иного события. Можно сказать, что одно событие наступает чаще, чем другое. В этом случае говорят, что оно более возможно, т. е. его наступление более вероятно. Естественно, при такой оценке человеку помогает здравый смысл и жизненный опыт. Например, из опыта известно, что при выполнении многих видов работ вредна торопливость. В спешке можно совершить такое действие, которое сведет на нет всю предыдущую работу. Иначе говоря, при спешке более вероятен брак в работе, т. е. вероятность (возможность) выхода брака выше.

Однако в жизни чаще встречаются события, сравнить или оценить возможности появления которых, основываясь на чисто интуитивных соображениях, трудно. Например, это можно сказать про события «герб появился два раза при пятикратном бросании монеты», «во время решения задачи отказала ЭВМ» и т. д. Как видно из приведенных примеров, каждое событие обладает определенной степенью возможности наступления, т. е. определенной оценкой. Такую оценку события называют вероятность события.

Вероятность события – это числовая характеристика возможности наступления случайного события в результате испытаний при заданной совокупности условий.

По определению, событию можно поставить в соответствие определенное число – его вероятность. Однако приведенное определение не дает формулу для нахождения этого числа.

Существует несколько подходов к определению вероятности. Рассмотрим основные из них.

1.2.1. Классическое определение вероятности

Классическое определение вероятности связано с определением благоприятствующего исхода.

Элементарные исходы, при которых данное событие наступает, называются *благоприятствующими* этому событию.

Пример 1.16. Бросается игральный кубик. Элементарные исходы «появление двух очков», «появление четырех очков», «появление шести очков» являются благоприятствующими событию «выпало четное число очков».

Вероятностью события A называется отношение числа m благоприятствующих событию A элементарных исходов к общему числу n равновозможных, единственно возможных и несовместных элементарных исходов:

$$P(A) = \frac{m}{n}.$$

Из определения вероятности события A следует, что $0 \le m \le n$, поэтому всегда выполняются неравенства $0 \le P(A) \le 1$, т. е. *вероятность любого события есть неотрицательное число, не превышающее единицы.*

Вероятность *невозможного* события равна нулю, т. е. $P(U) = 0$.

Вероятность *достоверного* события равна единице, т. е. $P(E) = 1$.

Равновозможные элементарные события являются *равновероятными*, т. е. обладают одной и той же вероятностью.

Тест 1.7. Вероятность события принимает любое значение из промежутка:

1) [0; 1];
2) [–1; 1];
3) $(-\infty; 1]$;
4) [0; 2];
5) $(0; +\infty)$.

Пример 1.17. Из ящика, в котором находится 4 белых, 6 желтых и 10 красных шаров, наугад вынимают один шар. Какова вероятность появления белого шара?

Решение

Обозначим через A событие «появление белого шара». В этом испытании элементарным исходом является извлечение из ящика любого одного шара. Число всех таких исходов равно числу шаров в ящике, т. е. $n = 20$. Число исходов, благоприятствующих появлению белого шара (событию A), равно числу белых шаров в урне, т. е. $m = 4$. Поэтому

$$P(A) = \frac{4}{20} = \frac{1}{5} = 0{,}2.$$

Ответ: 0,2.

Согласно классическому определению подсчет вероятности события A сводится к подсчету числа благоприятствующих ему исходов. Делают это обычно комбинаторными методами.

Комбинаторика – раздел математики, в котором изучаются вопросы о том, сколько различных комбинаций, подчиненных определенным условиям, можно составить из элементов, безразлично какой природы, заданного конечного множества.

Формулы комбинаторики используют при непосредственном вычислении вероятностей.

Перестановками называют комбинации, состоящие из одних и тех же различных элементов и отличающиеся друг от друга только их порядком. Число всевозможных перестановок из n элементов обозначается через P_n и вычисляется по формуле

$$P_n = n!,$$

где $n!$ (читается эн-факториал) = $1 \cdot 2 \cdot 3 \cdot \ldots \cdot n$, причем $1!=1$, $0!=1$.

Размещениями называют комбинации, составленные из n различных элементов по m элементов, которые отличаются либо составом элементов, либо их порядком. Число всех возможных размещений определяется формулой

$$A_n^m = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-m+1)$$

или

$$A_n^m = \frac{n!}{(n-m)!}.$$

Сочетаниями называются комбинации, составленные из n различных элементов по m элементов, которые отличаются хотя бы одним элементом. Число сочетаний из n элементов вычисляется по формуле

$$C_n^m = \frac{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-m+1)}{1 \cdot 2 \cdot 3 \cdot \ldots \cdot m}$$

или

$$C_n^m = \frac{n!}{m! \cdot (n-m)!}.$$

$$C_n^0 = C_n^n = 1.$$

Пример 1.18. Сколько трехзначных чисел можно составить из цифр 1, 2, 3, если каждая цифра входит в запись числа только один раз?

Решение

Число различных комбинаций из трех цифр равно 3!. В результате число различных трехзначных чисел равно $P_3 = 3! = 1 \cdot 2 \cdot 3 = 6$.

Ответ: 6.

Пример 1.19. Правление коммерческого банка выбирает из восьми кандидатов три человека на различные должности (все 8 кандидатов имеют равные шансы). Сколькими способами это можно сделать?

Решение

Так как группы по 3 человека могут отличаться и составом претендентов, и порядком заполнения ими трех вакансий, то для ответа необходимо рассчитать число размещений из 8 элементов по 3, т. е.

$$A_8^3 = 8 \cdot 7 \cdot 6 = 336$$

или

$$A_8^3 = \frac{8!}{(8-3)!} = \frac{8!}{5!} = \frac{5! \cdot 6 \cdot 7 \cdot 8}{5!} = 6 \cdot 7 \cdot 8 = 336.$$

(Так как $8! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8$; $5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5$, то 8! можем представить через 5!, т. е. $8! = 5! \cdot 6 \cdot 7 \cdot 8$).

Ответ: 336.

Пример 1.20. В шахматном турнире участвуют 16 человек. Сколько партий должно быть сыграно в турнире, если между любыми двумя участниками должна быть сыграна одна партия?

Решение

Каждая партия играется двумя участниками из 16 и отличается от других только составом пар участников, т. е. представляет собой сочетание из 16 элементов по 2. Их число находим по формуле

$$C_{16}^2 = \frac{16 \cdot 15}{1 \cdot 2} = 120,$$

или

$$C_{16}^2 = \frac{16!}{2! \cdot (16-2)!} = \frac{16!}{2! \cdot 14!} = \frac{14! \cdot 15 \cdot 16}{2! \cdot 14!} = \frac{15 \cdot 16}{2!} = \frac{15 \cdot 16}{1 \cdot 2} = 120.$$

Ответ: 120.

Тест 1.8. Порядок выступления 7 участников конкурса определяется жребием. Число различных вариантов жеребьевки при этом равно:

1) $\frac{1}{7!}$;

2) P_7;

3) C_7^1;

4) $\frac{1}{P_7}$;

5) A_7^1.

Многие комбинаторные задачи могут быть решены с помощью следующих двух важных правил, называемых соответственно правилами суммы и произведения.

Правило суммы. Если элемент A_1 может быть выбран n_1 способами, а элемент A_2 может быть выбран n_2 способами, то выбор одного из элементов: или A_1, или A_2 может быть осуществлен $n_1 + n_2$ способами.

Правило произведения. Если элемент A_1 может быть выбран n_1 способами, после каждого такого выбора элемент A_2 может быть выбран n_2 способами, то выбор всех элементов A_1, A_2 в указанном порядке может быть осуществлен $n_1 \cdot n_2$ способами.

Пример 1.21. Сколько чисел, содержащих не менее трех попарно различных цифр, можно составить из цифр 2, 4, 6, 8, 9?

Решение

По правилу произведения трехзначных чисел можно составить $5 \cdot 4 \cdot 3 = 60$, а четырехзначных – $5 \cdot 4 \cdot 3 \cdot 2 = 120$, столько же пятизначных чисел ($5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$). По правилу суммы, всего можно составить $60 + 120 + 120 = 300$ чисел, состоящих не менее чем из трех попарно различных цифр.

Пример 1.22. Дано шесть карточек с буквами Н, М, И, Я, Л, О. Какова вероятность того, что получится слово «МОЛНИЯ», если наугад одна за другой выбираются шесть карточек и располагаются в ряд в порядке появления?

Решение

Шестибуквенные слова отличаются друг от друга лишь порядком расположения букв («НОЛМИЯ», «ЯНОЛИМ», «ОЛНИЯМ» и т. д.). Их число равно числу перестановок из 6 букв, т. е. $n = P_6$. Событие A, состоящее в том, что получится слово «МОЛНИЯ», очевидно, появляется лишь в одном элементарном исходе, т. е. $m = 1$. Следовательно,

$$P(A) = \frac{m}{n} = \frac{1}{P_6} = \frac{1}{6!} = \frac{1}{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6} = \frac{1}{720} \approx 0{,}0014.$$

Ответ: $\approx 0{,}0014$.

Тест 1.9. Буквы Т, Е, И, Я, Р, О написаны на отдельных карточках. Ребенок берет карточки в случайном порядке и прикладывает одну к другой все 6 карточек. Вероятность того, что получится слово «ТЕОРИЯ», равна:

1) $\frac{1}{A_6^6}$;

2) $\frac{6}{P_5}$;

3) $\frac{1}{C_6^6}$;

4) $\frac{6}{6!}$;

5) $\frac{1}{P_6}$.

Пример 1.23. Из 30 студентов 10 имеют спортивные разряды. Какова вероятность того, что выбранные наудачу 3 студента имеют разряд?

Решение

Пусть событие A – «3 выбранных наудачу студента имеют разряд». Общее число всех возможных исходов из 30 студентов по 3 равно $n = C_{30}^3$, так как комбинации из 30 студентов по 3 представляют собой сочетания, поскольку отличаются хотя бы одним студентом. Точно также число исходов, благоприятствующих событию A, равно C_{10}^3. Итак,

$$P(A) = \frac{m}{n} = \frac{C_{10}^3}{C_{30}^3} = \frac{10!}{3!\cdot(10-3)!} : \frac{30!}{3!\cdot(30-3)!} = \frac{10!}{3!\cdot 7!}\cdot\frac{3!\cdot 27!}{30!} =$$

$$= \frac{7!\cdot 8\cdot 9\cdot 10}{3!\cdot 7!}\cdot\frac{3!\cdot 27!}{27!\cdot 28\cdot 29\cdot 30} = \frac{8\cdot 9\cdot 10}{28\cdot 29\cdot 30} = \frac{6}{203} \approx 0{,}03.$$

Ответ: $\approx 0{,}03$.

Тест 1.10. В ящике 30 деталей, среди которых 20 окрашенных. Сборщик наудачу извлекает три детали. Вероятность того, что извлеченные детали окажутся окрашенными, равна:

1) $\frac{C_{10}^3}{C_{30}^{10}}$;

2) $\frac{1}{C_{30}^3}$;

3) $\frac{C_{10}^3}{C_{20}^3}$;

4) $\frac{C_{20}^3}{C_{30}^3}$;

5) $\frac{C_{30}^3}{C_{10}^3}$.

Пример 1.24. Подбрасывают два игральных кубика. Какова вероятность того, что сумма выпавших очков не превосходит 7?

Решение

Каждый из кубиков может упасть шестью различными способами. Тогда два кубика по правилу произведения могут упасть $6 \cdot 6 = 36$ различными способами. Каждому такому способу соответствует событие, которое является исходом испытания «подбрасывание двух кубиков». В силу симметричности кубиков все эти события равновозможны и образуют полную группу несовместных событий. Поэтому число всех исходов подбрасывания двух кубиков $n = 6 \cdot 6 = 36$. Обозначим через B событие «сумма выпавших очков не превосходит 7». Событию B благоприятствуют 21 элементарный исход: (1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (3,1), (4,1), (5,1), (6,1), (2,2), (2,3), (2,4), (2,5), (3,2), (4,2), (5,2), (3,3), (3,4), (4,3). Следовательно, $m = 21$. Тогда

$$P(B) = \frac{21}{36} = \frac{7}{12} \approx 0{,}58.$$

Ответ: $\approx 0{,}58$.

Тест 1.11. Подбрасывают три монеты. Вероятность того, что при этом (безразлично в каком порядке) выпадет два раза «герб» и один раз «надпись», равна:

1) $\frac{5}{6}$;

2) $\frac{3}{8}$;

3) $\frac{1}{2}$;

4) $\frac{2}{8}$;

5) $\frac{1}{6}$.

1.2.2. Геометрическое определение вероятности

Чтобы преодолеть недостаток классического определения вероятности, состоящий в том, что оно неприменимо к испытаниям с бесконечным числом исходов, вводят геометрические вероятности – вероятности попадания точки в область (отрезок, часть плоскости и т.д.).

Пусть, например, плоская фигура g составляет часть плоской фигуры G. На фигуру G наудачу бросается точка. Это означает, что все точки области G «равноправны» в отношении попадания туда брошенной случайной точки. Полагая, что вероятность события A – попадания брошенной точки на фигуру g – пропорциональна площади этой фигуры и не зависит ни от ее расположения относительно G, ни от формы g (рис. 1), найдем:

$$P(A) = \frac{S_g}{S_G},$$

где S_g и S_G – соответственно площади областей g и G.

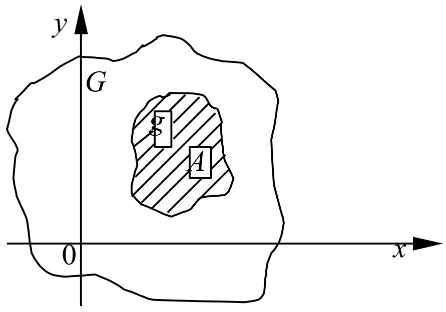

Рис. 1

Фигуру g называют *благоприятствующей событию A*.

Область, на которую распространяется понятие геометрической вероятности, может быть одномерной (прямая, отрезок) и трехмерной (некоторое тело в пространстве). Обозначая меру (длину, площадь, объем) области через *mes*, приходим к следующему определению.

Геометрической вероятностью события A называется отношение меры области, благоприятствующей появлению события A, к мере всей области:

$$P(A) = \frac{mesg}{mesG}.$$

Пример 1.25. В круг вписан квадрат. Какова вероятность того, что точка, наудачу поставленная в круге, окажется внутри квадрата?

Решение

Площадь круга $S = \pi r^2$, площадь квадрата $S_0 = 2r^2$, где r – радиус круга, A – «попадание точки в квадрат» (рис. 2). Отсюда по определению, полагая $S_G + S$, $S_g = S_0$, находим искомую вероятность:

$$P(A) = \frac{S_0}{S} = \frac{2r^2}{\pi r^2} = \frac{2}{\pi} \approx 0{,}637.$$

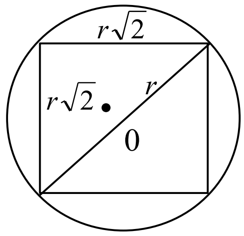

Рис. 2

Ответ: $\approx 0{,}637$.

Тест 1.12. Два лица A и B условились встретиться в определенном месте, договорившись только о том, что каждый является туда в любой момент времени между 11 и 12 часами и ждет в течение 30 минут. Если партнер к этому времени еще не пришел или уже успел покинуть установленное место, встреча не состоится. Для нахождения вероятности того, что встреча состоится необходимо использовать:

1) статистическое определение вероятности;

2) геометрическое определение вероятности.

1.2.3. Статистическое определение вероятности

Чтобы дать это определение, предварительно вводят понятие относительной частоты события.

Относительной частотой события, или *частотой,* называется отношение числа испытаний, в которых появилось это событие, к числу всех произведенных испытаний. Обозначим частоту события A через $W(A)$, тогда по определению

$$W(A) = \frac{m}{n},$$

где m – число испытаний, в которых появилось событие A;

n – число всех произведенных испытаний.

Частота события обладает следующими *свойствами:*

1. Частота случайного события есть число, заключенное между нулем и единицей:

$$0 < W(A) < 1.$$

2. Частота достоверного события E равна единице:

$$W(E) = 1.$$

3. Частота невозможного события равна нулю:

$$W(U) = 0.$$

4. Частота суммы двух несовместных событий A и B равна сумме частот этих событий:

$$W(A + B) = W(A) + W(B).$$

Наблюдения позволили установить, что относительная частота обладает свойствами статистической устойчивости: в различных сериях многочленных испытаний (в каждом из которых может появиться или не появиться это событие) она принимает значения, достаточно близкие к некоторой постоянной. Эту постоянную, являющуюся объективной числовой характеристикой явления, считают вероятностью данного события.

Статистической вероятностью события называется число, около которого колеблется относительная частота события при достаточно большом числе испытаний.

Пример 1.26. Игральный кубик подброшен 60 раз, при этом шестерка появилась 10 раз. Какова частота появления шестерки?

Решение

Из условия задачи следует, что $n = 60$, $m = 10$, поэтому

$$W = \frac{10}{60} = \frac{1}{6} \approx 0{,}17.$$

Ответ: $\approx 0{,}17$.

Пример 1.27. При стрельбе из винтовки относительная частота попадания в цель оказалась равной 0,85. Каково число попаданий, если всего было произведено 120 выстрелов?

Решение

Из формулы $W = \dfrac{m}{n}$ следует, что $m = W \cdot n$. Так как $W = 0{,}85$, $n = 120$, то $m = 0{,}85 \cdot 120 = 102$. Таким образом, было получено 102 попадания.

Ответ: 102.

1.3. Теоремы сложения и умножения вероятностей

Теорема сложения вероятностей двух совместных событий

Вероятность суммы двух совместных событий равна сумме вероятностей этих событий без вероятности их совместного появления:

$$P(A+B) = P(A) + P(B) - P(AB).$$

Теорема сложения вероятностей двух несовместных событий

Вероятность суммы двух несовместных событий равна сумме вероятностей этих событий:

$$P(A+B) = P(A) + P(B).$$

Теорема сложения вероятностей n несовместных событий

Вероятность суммы n несовместных событий $A_1, A_2, \ldots, A_n$ равна сумме вероятностей этих событий:

$$P(A_1 + A_2 + \ldots + A_n) = P(A_1) + P(A_2) + \ldots + P(A_n).$$

Сумма вероятностей противоположных событий равна единице:

$$P(A) + P(\overline{A}) = 1.$$

Если обозначить $P(A) = p$, $P(\overline{A}) = q$, то формула примет вид:

$$p + q = 1.$$

Перед тем как излагать теорему умножения вероятностей, введем еще одно важное понятие – понятие о независимых и зависимых событиях.

Событие A называется *независимым* от события B, если вероятность события A не зависит от того, произошло событие B или нет.

Событие A называется *зависимым* от события B, если вероятность события A меняется в зависимости от того, произошло событие B или нет.

Пример 1.28. Опыт состоит в подбрасывании двух монет. Рассматриваются события: A – «появление герба на первой монете», B – «появление герба на второй монете». В данном случае вероятность события A не зависит от того, произошло событие B или нет, событие A независимо от события B.

Пример 1.29. В ящике два белых шара и один черный, два лица вынимают из ящика по одному шару. Рассматриваются события: A – «появление белого шара у 1-го лица», B – «появление белого шара у 2-го лица». Вероятность события A до того, как известно что-либо о событии B, равна $\dfrac{2}{3}$. Если стало известно, что событие B произошло, то вероятность события A становится равной $\frac{1}{2}$, из чего заключаем, что событие A зависит от события B.

Вероятность события B при условии, что произошло событие A, называется *условной вероятностью* события B и обозначается $P_A(B)$.

Теорема умножения вероятностей двух зависимых событий

Вероятность произведения двух зависимых событий равна произведению вероятности одного из них на условную вероятность другого при условии, что первое событие произошло:

$$P(AB) = P(A) \cdot P_A(B)$$

или

$$P(AB) = P(B) \cdot P_B(A).$$

Условная вероятность исчисляется по следующей формуле:

$$P_B(A) = \frac{P(AB)}{P(B)}, \ P(B) \neq 0.$$

Теорема умножения вероятностей двух независимых событий

Вероятность произведения двух независимых событий равна произведению их вероятностей:

$$P(AB) = P(A) \cdot P(B).$$

Теорема умножения вероятностей n зависимых событий

Вероятность произведения n зависимых событий равна произведению одного из них на условные вероятности всех остальных, вычисленные в предположении, что все предыдущие события наступят:

$$P(A_1 \cdot A_2 \cdot \ldots \cdot A_n) = P(A_1) \cdot P_{A_1}(A_2) \cdot P_{A_1 A_2}(A_3) \cdot \ldots \cdot P_{A_1 A_2 \ldots A_{n-1}}(A_n).$$

События $A_1, A_2, \ldots, A_n$ называются независимыми в совокупности, если вероятность наступления любого из них не зависит от того, наступила или нет любая комбинация остальных.

Теорема умножения вероятностей n независимых событий

Если события $A_1, A_2, \ldots, A_n$ независимы в совокупности, то вероятность их произведения равна произведению вероятностей этих событий:

$$P(A_1 \cdot A_2 \cdot \ldots \cdot A_n) = P(A_1) \cdot P(A_2) \cdot \ldots \cdot P(A_n).$$

Вычисление вероятности суммы событий можно свести к вычислению вероятности произведения противоположных событий по формуле

$$P(A_1 + A_2 + \ldots + A_n) = 1 - P(\overline{A_1} \cdot \overline{A_2} \cdot \ldots \cdot \overline{A_n}) =$$
$$= 1 - P(\overline{A_1}) \cdot P(\overline{A_2}) \cdot \ldots \cdot P(\overline{A_n})$$

или

$$P(A_1 + A_2 + \ldots + A_n) = 1 - q_1 \cdot q_2 \cdot \ldots \cdot q_n.$$

Если независимые события $A_1, A_2, \ldots, A_n$ имеют одинаковую вероятность, равную p, то вероятность появления хотя бы одного из этих событий выражается формулой

$$P(A_1 + A_2 + \ldots + A_n) = 1 - q^n.$$

Пример 1.30. В ящике 12 зеленых, 7 черных и 11 синих шаров. Наудачу вынимается один шар. Какова вероятность того, что вынутый шар не зеленый?

Решение

Если вынутый шар не зеленый, то это означает, что он либо черный, либо синий. Пусть событие A – «появился черный шар», а событие B – «появился синий шар». Найдем вероятности этих событий:

$$P(A) = \frac{7}{12+7+11} = \frac{7}{30}, \ P(B) = \frac{11}{12+7+11} = \frac{11}{30}.$$

Так как события A и B несовместны, то

$$P(A+B)=P(A)+P(B)=\frac{7}{30}+\frac{11}{30}=\frac{18}{30}=\frac{3}{5}=0{,}6\,.$$

Ответ: 0,6.

Пример 1.31. Подбрасывается игральный кубик. Чему равна вероятность того, что выпадет четное число очков?

Решение

Пусть событие A – «выпало четное число очков», событие B_1 – «выпало 2 очка», событие B_2 – «выпало 4 очка», событие B_3 – «выпало 6 очков». Событие A означает, что наступило хотя бы одно из событий: B_1, B_2, B_3, т. е. $A = B_1 + B_2 + B_3$. Поскольку события B_1, B_2, B_3 несовместны, то

$$P(A)=P(B_1)+P(B_2)+P(B_3)=\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{1}{2}=0{,}5\,.$$

Ответ: 0,5.

Пример 1.32. В ящике 5 белых и 2 черных шара. Из него извлекают наугад 3 шара. Какова вероятность того, что при первом извлечении появится белый шар, при втором – снова белый шар и при третьем – черный?

Решение

Пусть событие A_1 – «первый шар белый», событие A_2 – «второй шар белый», событие A_3 – «третий шар черный». Нас интересует появление и события A_1, и события A_2, и события A_3, т. е. их произведения $A_1 \cdot A_2 \cdot A_3$. События A_1, A_2 и A_3 зависимы, так как наступление события A_1 влияет на вероятность события A_2 (шаров в ящике останется 6, из них только 4 белых), наступление событий A_1 и A_2 влияет на вероятность события A_3 (шаров в ящике останется 5). Поэтому

$$P(A_1\cdot A_2\cdot A_3)=P(A_1)\cdot P_{A_1}(A_2)\cdot P_{A_1A_2}(A_3)=\frac{5}{7}\cdot\frac{4}{6}\cdot\frac{2}{5}=\frac{4}{21}\approx 0{,}1905.$$

Ответ: $\approx 0{,}1905$.

Тест 1.13. Кулинар изготовил 15 омлетов, причем 4 пересолил. Вероятность того, что из 3 случайно выбранных омлетов все окажутся непересоленными, равна:

1) $\frac{C_4^0}{C_{15}^3}$;

2) $\frac{C_{11}^3}{C_{15}^4}$;

3) $\frac{C_4^0\cdot C_{11}^3}{C_{15}^3}$;

4) $\frac{C_{11}^4}{C_{15}^4}$;

5) $\frac{4}{15}$.

Тест 1.14. Из колоды в 36 карт наугад одна за другой извлекаются две карты. Вероятность того, что ими оказались два короля, равна:

1) $\frac{1}{9}\cdot\frac{3}{35}$;

2) $\frac{1}{4}\cdot\frac{8}{35}$;

3) $\frac{6}{36}\cdot\frac{3}{35}$;

4) $\frac{4}{36}\cdot\frac{4}{35}$;

5) $\frac{1}{9}\cdot\frac{4}{36}$.

Тест 1.15. В ящике находятся 6 черных, 5 красных и 4 белых шара. Последовательно вынимают три шара. Вероятность того, что первый шар окажется черным, второй – красным и третий – белым, равна:

1) $\frac{6}{15}\cdot\frac{5}{15}\cdot\frac{4}{15}$;

2) $\frac{1}{15}\cdot\frac{5}{15}\cdot\frac{4}{15}$;

3) $\frac{15}{C_{15}^3}$;

4) $\frac{2}{5}\cdot\frac{5}{14}\cdot\frac{4}{15}$;

5) $\frac{2}{5}\cdot\frac{5}{14}\cdot\frac{4}{13}$.

Пример 1.33. В ящике находятся 5 белых и 2 черных шара. Из него извлекают наугад 3 шара, каждый раз шары возвращаются в ящик и перемешиваются. Какова вероятность того, что при первом извлечении появится белый шар, при втором – снова белый шар и при третьем – черный?

Решение

Из решения примера 1.32 следует, что если после первого и второго извлечения шары возвращаются в ящик, то события A_1, A_2 и A_3 – независимы, откуда

$$P(A_1\cdot A_2\cdot A_3)=P(A_1)\cdot P(A_2)\cdot P(A_3)=\frac{5}{7}\cdot\frac{5}{7}\cdot\frac{2}{7}=\frac{50}{343}\approx 0{,}1458.$$

Ответ: $\approx 0{,}1458$.

Тест 1.16. Вероятность поломки первого станка в течение смены равна 0,2, а второго – 0,13. Вероятность того, что оба станка потребуют наладки в течение смены, равна:

1) 0,33;
2) 0,026;
3) 0,26;
4) 0,22;
5) 0,43.

Пример 1.34. Из букв А, С, Н, Н, А, А разрезной азбуки составляется наудачу слово, состоящее из 6 букв. Какова вероятность того, что получится слово «АНАНАС»?

Решение

Пусть событие A – «получится слово АНАНАС». Обозначим через A_1, A_2, A_3, A_4, A_5, A_6, события «появления букв» А, Н, А, Н, А, С соответственно. События A_i (I = 1,2,…,6) – зависимые. Нас интересует появление события $A = A_1\cdot A_2\cdot A_3\cdot A_4\cdot A_5\cdot A_6$. Поэтому

$$P(A)=P(A_1\cdot A_2\cdot A_3\cdot A_4\cdot A_5\cdot A_6)=P(A_1)\cdot P_{A_1}(A_2)\cdot P_{A_1A_2}(A_3)\times$$
$$\times P_{A_1A_2A_3}(A_4)\cdot P_{A_1A_2A_3A_4}(A_5)\cdot P_{A_1A_2A_3A_4A_5}(A_6)=$$
$$=\frac{3}{6}\cdot\frac{2}{5}\cdot\frac{2}{4}\cdot\frac{1}{3}\cdot\frac{1}{2}\cdot\frac{1}{1}=\frac{1}{60}\approx 0{,}017.$$

Ответ: $\approx 0{,}017$.

Пример 1.35. Три стрелка стреляют в цель независимо друг от друга. Вероятность попадания в цель для первого стрелка равна 0,6, для второго – 0,7 и для третьего – 0,8. Найти вероятность того, что при трех выстрелах произойдет хотя бы одно попадание в цель.

Решение

Обозначим через A событие «при трех выстрелах произойдет хотя бы одно попадание в цель», и через $\overline{A}$ – противоположное ему событие, т. е. «при каждом из трех выстрелов попадания в цель не произойдет». Через B, C, D обозначим «попадание в цель соответственно первым стрелком, вторым и третьим», через $\overline{B}$, $\overline{C}$, $\overline{D}$ – соответственно «промахи первым, вторым и третьим стрелками». События *B, C, D* – независимые, поэтому

$$P(A) = 1 - P(\overline{B}) \cdot P(\overline{C}) \cdot P(\overline{D}).$$

Так как $P(\overline{B})$ = 1 – $P(B)$ = 1 – 0,6 = 0,4; $P(\overline{C})$ = 1 – $P(C)$ = 1 – 0,7 = 0,3; $P(\overline{D}) = 1 - P(D) = 1 - 0{,}8 = 0{,}2$, то $P(A) = 1 - 0{,}4 \cdot 0{,}3 \cdot 0{,}2 = 0{,}976$.

Ответ: $\approx 0{,}976$.

Пример 1.36. Вероятность выигрыша по одному лотерейному билету составляет p = 0,03. Некто имеет 7 билетов. Какова вероятность выигрыша хотя бы по одному билету?

Решение

Обозначим событие «выигрыш хотя бы по одному билету» буквой *A*. Тогда $P(A) = 1 - q^7$, так как $q = 1 - p$ = 1 – 0,03 = 0,97, то $P(A) = 1 - (0{,}97)^7 \approx 0{,}192$.

Ответ: $\approx 0{,}192$.

Пример 1.37. Подбрасывают два игральных кубика. Какова вероятность появления хотя бы один раз шести очков?

Решение

Пусть событие *A* – «появление шести очков на первом кубике», событие *B* – «появление шести очков на втором кубике». Тогда *A* + *B* – «появление хотя бы один раз шести очков при подбрасывании игральных кубиков». События *A* и *B* совместные, поэтому применим формулу

$$P(A+B) = P(A) + P(B) - P(AB).$$

Поскольку A и B – независимые события, то $P(AB) = P(A) \cdot P(B)$. Найдем вероятности этих событий:

$$P(A) = \frac{1}{6},\ P(B) = \frac{1}{6}.$$

Итак, $P(A+B) = \frac{1}{6} + \frac{1}{6} - \frac{1}{6} \cdot \frac{1}{6} = \frac{11}{36} \approx 0{,}306$.

Ответ: $\approx 0{,}306$.

Тест 1.17. Вероятность того, что потребитель увидит рекламу определенного продукта по телевидению, равна 0,04. Вероятность того, что потребитель увидит рекламу того же продукта на рекламном стенде, равна 0,06. Предполагается, что оба события независимы. Вероятность того, что потребитель увидит хотя бы одну рекламу, равна:

1) 0,04 + 0,06;
2) 0,04 · 0,06;
3) 0,04 – 0,04 · 0,06;
4) 0,04 + 0,06 – 0,0024;
5) $\frac{0{,}04 \cdot 0{,}06}{0{,}0024}$.

1.4. Формула полной вероятности. Формула Байеса

Пусть событие *A* может наступить при условии появления одного из несовместных событий B_1, B_2, …, B_n, образующих полную группу. Так как заранее не известно, какое из этих событий наступит, их называют *гипотезами.* Тогда вероятность события *A* вычисляется по *формуле полной вероятности:*

$$P(A) = P(B_1) \cdot P_{B_1}(A) + P(B_2) \cdot P_{B_2}(A) + \ldots + P(B_n) \cdot P_{B_n}(A)$$

или

$$P(A) = \sum_{i=1}^{n} P(B_i) \cdot P_{B_i}(A),$$

где $P(B_i)$ – вероятность гипотезы B_i;

$P_{B_i}(A)$ – *условная вероятность события A* при этой гипотезе.

Если до испытания вероятности гипотез были $P(B_1)$, $P(B_2)$, …, $P(B_n)$, а в результате испытания появилось событие A, то с учетом этого события «новые», т. е. условные, вероятности гипотез вычисляются по *формуле Байеса:*

$$P_A(B_i) = \frac{P(B_i)\cdot P_{B_i}(A)}{\sum_{i=1}^{n} P(B_i)\cdot P_{B_i}(A)} = \frac{P(B_i)\cdot P_{B_i}(A)}{P(A)},\ i = 1,2,\dots,n.$$

Формула Байеса дает возможность произвести переоценки вероятностей гипотез B_1, B_2, …, B_n после того, как стало известно, что событие A наступило.

Тест 1.18. Вероятность того, что клиент банка не вернет заем в период экономического роста, равна 0,04 и 0,13 – в период экономического кризиса. Предположим, что вероятность того, что начнется период экономического роста, равна 0,65. Количество гипотез события «случайно выбранный клиент банка не вернет полученный кредит», равно:

1) 2;
2) 4;
3) 0;
4) 1;
5) 3.

Пример 1.38. Электролампы изготовляются на двух заводах. Первый завод производит 60% общего количества электроламп, второй – 40%. Продукция первого завода содержит 70% стандартных ламп, второго – 80%. В магазин поступает продукция обоих заводов. Какова вероятность того, что купленная в магазине лампа окажется стандартной?

Решение

Обозначим через A событие, состоящее в том, что куплена стандартная лампа, B_1 – «лампа изготовлена на первом заводе», B_2 – «лампа изготовлена на втором заводе».

Из условия следует, что

$$P(B_1) = \frac{60}{100} = 0{,}6;\quad P(B_2) = \frac{40}{100} = 0{,}4;$$

$$P_{B_1}(A) = \frac{70}{100} = 0{,}7;\quad P_{B_2}(A) = \frac{80}{100} = 0{,}8.$$

По формуле полной вероятности имеем:

$$P(A) = P(B_1)\cdot P_{B_1}(A) + P(B_2)\cdot P_{B_2}(A) = 0{,}6\cdot 0{,}7 + 0{,}4\cdot 0{,}8 = 0{,}74.$$

Ответ: 0,74.

Тест 1.19. Издательство разослало рекламные материалы на новый учебник по бухгалтерскому учету, которые получили 80% профессоров, читающих этот курс в различных высших учебных заведениях. Отобрали эту книгу и приняли ее для преподавания 30% профессоров, получивших рекламные материалы и 10% не получивших их. Вероятность того, что случайно выбранный профессор вуза принял этот учебник для преподавания, равна:

1) 0,8 + 0,3 + 0,1;
2) $\dfrac{0{,}8\cdot 0{,}3}{0{,}8\cdot 0{,}3 + 0{,}2\cdot 0{,}1}$;
3) 1,1 · 0,1;
4) 0,8 · 0,1 + 0,3 · 0,1;
5) 0,8 · 0,3 + 0,2 · 0,1.

Пример 1.39. В студенческой группе 70% – юноши, 20% юношей и 40% девушек имеют сотовый телефон. После занятий в аудитории был найден кем-то забытый телефон. Какова вероятность того, что он принадлежал юноше?

Решение

Пусть событие A состоит в том, что после занятий в аудитории был найден кем-то забытый телефон. Тогда в качестве гипотез примем события B_1 – «найденный телефон принадлежал юноше» и B_2 – «найденный телефон принадлежал девушке». Из условия следует, что

$$P(B_1)=\frac{70}{100}=0{,}7\,;\quad P(B_2)=1-P(B_1)=1-0{,}7=0{,}3\,;$$

$$P_{B_1}(A)=\frac{20}{100}=0{,}2\,;\quad P_{B_2}(A)=\frac{40}{100}=0{,}4\,.$$

В соответствии с формулой Байеса находим искомую вероятность:

$$P_A(B_1)=\frac{P(B_1)\cdot P_{B_1}(A)}{P(B_1)\cdot P_{B_1}(A)+P(B_2)\cdot P_{B_2}(A)}=\frac{0{,}7\cdot 0{,}2}{0{,}7\cdot 0{,}2+0{,}3\cdot 0{,}4}=\frac{0{,}14}{0{,}26}\approx 0{,}54.$$

Ответ: $\approx 0{,}54$.

Тест 1.20. Медицинский тест на возможность вирусного заболевания дает следующие результаты:
1. Если проверяемый болен, то тест даст положительный результат с вероятностью 0,92.
2. Если проверяемый не болен, то тест может дать положительный результат с вероятностью 0,04.

Поскольку заболевание редкое, то ему подвержены только 0,1% населения. Предположим, что некоторому случайно выбранному человеку сделан анализ и получен положительный результат. Вероятность того, что человек действительно болен, равна:

1) $\dfrac{0{,}001\cdot 0{,}92}{0{,}001\cdot 0{,}92+0{,}999\cdot 0{,}04}$;

2) $0{,}001\cdot 0{,}92+0{,}999\cdot 0{,}04$;

3) 0,001 + 0,04;

4) $\dfrac{0{,}999\cdot 0{,}04}{0{,}001\cdot 0{,}92+0{,}999\cdot 0{,}04}$;

5) $0{,}001\cdot 0{,}04+0{,}92\cdot 0{,}999$.

Пример 1.40. В коробке 3 новых и 3 уже использованных теннисных мяча. Для первой игры наудачу берут из коробки 2 мяча и затем их возвращают в коробку. Какова вероятность для второй игры из этой коробки наудачу вынуть два новых мяча?

Решение

Введем событие A – «вынуть два новых мяча для второй игры». Ситуация перед второй игрой описывается следующими взаимоисключающими возможностями:
- B_1 – «в коробке 1 новый мяч», если в первый раз играли двумя новыми мячами;
- B_2 – «в коробке 2 новых мяча», если играли одним старым и одним новым мячами;
- B_3 – «в коробке 3 новых мяча», если играли двумя старыми мячами.

События B_1, B_2, B_3 образуют полную группу несовместных событий, при этом:

$$P(B_1)=\frac{C_3^2}{C_6^2}=\frac{\dfrac{3!}{2!\cdot(3-2)!}}{\dfrac{6!}{2!\cdot(6-2)!}}=\frac{\dfrac{2!\cdot 3}{2!\cdot 1!}}{\dfrac{4!\cdot 5\cdot 6}{2!\cdot 4!}}=\frac{3}{15}=\frac{1}{5};$$

$$P(B_2)=\frac{C_3^1\cdot C_3^1}{C_6^2}=\frac{3\cdot 3}{\dfrac{6!}{2!\cdot(6-2)!}}=\frac{9}{\dfrac{4!\cdot 5\cdot 6}{2!\cdot 4!}}=\frac{9}{15}=\frac{3}{5};\ P(B_3)=\frac{C_3^2}{C_6^2}=\frac{\dfrac{3!}{2!\cdot(3-2)!}}{\dfrac{6!}{2!\cdot(6-2)!}}=\frac{\dfrac{2!\cdot 3}{2!\cdot 1!}}{\dfrac{4!\cdot 5\cdot 6}{2!\cdot 4!}}=\frac{3}{15}=\frac{1}{5}.$$

Для контроля можно найти сумму вероятностей гипотез, она должна равняться единице: $\sum_{i=1}^{n}P(B_i)=\frac{1}{5}+\frac{3}{5}+\frac{1}{5}=1$.

Далее имеем: $P_{B_1}(A)=0$; $P_{B_2}(A)=\dfrac{C_2^2}{C_6^2}=\dfrac{1}{15}$; $P_{B_3}(A)=\dfrac{C_3^2}{C_6^2}=\dfrac{1}{5}$.

Теперь, используя формулу полной вероятности, найдем:

$$P(A) = P(B_1) \cdot P_{B_1}(A) + P(B_2) \cdot P_{B_2}(A) + P(B_3) \cdot P_{B_3}(A) =$$

$$= \frac{1}{5} \cdot 0 + \frac{3}{5} \cdot \frac{1}{15} + \frac{1}{5} \cdot \frac{1}{5} = \frac{2}{25} = 0{,}08.$$

Ответ: 0,08.

Тест 1.21. На химическом заводе установлена система аварийной сигнализации. Когда возникает аварийная ситуация, звуковой сигнал срабатывает с вероятностью 0,95. Звуковой сигнал может сработать случайно и без аварийной ситуации с вероятностью 0,02. Реальная вероятность аварийной ситуации равна 0,004. Предположим, что звуковой сигнал сработал. Вероятность реальной аварийной ситуации находится с помощью формулы:

1) полной вероятности;
2) Байеса.

Вопросы для самоконтроля

1. Что называют испытанием?
2. Что называют событием?
3. Какое событие называют достоверным в данном испытании?
4. Какое событие называют невозможным в данном испытании?
5. Какое событие называют случайным в данном испытании?
6. Что называют элементарным событием?
7. Какие события называют совместными в данном испытании?
8. Какие события называют несовместными в данном испытании?
9. Какие события образуют полную группу событий?
10. Какие события считают равновозможными?
11. Какие события называют противоположными?
12. Что называют суммой двух событий?
13. Что называют произведением двух событий?
14. Что называют перестановками?
15. По какой формуле вычисляют число перестановок из n различных элементов?
16. Что называют размещениями?
17. По какой формуле вычисляют число размещений из n различных элементов по m элементов?
18. Что называют сочетаниями?
19. По какой формуле вычисляют число сочетаний из n элементов по m элементов?
20. Что называют вероятностью события?
21. Чему равна вероятность достоверного события?
22. Чему равна вероятность невозможного события?
23. В каких пределах заключена вероятность любого события?
24. Что такое частота события?
25. Какое определение вероятности называют статистическим?
26. Какими свойствами обладает статистическая вероятность?
27. Как определяется геометрическая вероятность в общем случае?
28. Чему равна вероятность суммы двух совместных событий?
29. Чему равна вероятность суммы двух несовместных событий?
30. Как формируется теорема о вероятности суммы n несовместных событий.
31. Чему равна вероятность произведения двух независимых событий?
32. Чему равна вероятность произведения двух зависимых событий?
33. Что такое условная вероятность?
34. Как найти вероятность появления хотя бы одного из n независимых событий, имеющих одинаковые вероятности?
35. Какие события называются гипотезами?
36. Какие события образуют полную группу событий?
37. При каких условиях применяется формула полной вероятности?
38. Как изменяются вероятности гипотез, если известно, что событие произошло?

Ответы на тестовые задания

Номер теста	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	1.10
Правильный ответ	2	3	3	1	4	1	1	2	5	4

Номер теста	1.11	1.12	1.13	1.14	1.15	1.16	1.17	1.18	1.19	1.20	1.21
Правильный ответ	2	2	3	1	5	2	4	1	5	1	2

2. ПОВТОРНЫЕ НЕЗАВИСИМЫЕ ИСПЫТАНИЯ

2.1. Формула Бернулли

Пусть производятся n независимых испытаний, в каждом из которых может произойти некоторое событие A с одной и той же вероятностью $P(A) = p$ или произойти противоположное событие $\overline{A}$ с вероятностью $P(\overline{A}) = q = 1 - p$. Тогда вероятность того, что событие A наступит ровно k раз, находится по *формуле Бернулли*:

$$P_n(k) = C_n^k \cdot p^k \cdot q^{n-k}, \quad k = 0,1,2,...,n.$$

Пример 2.1. Игральный кубик подбрасывают 10 раз. Какова вероятность того, что шесть очков выпадет два раза?

Решение

По условию $n = 10$, $k = 2$, $p = \frac{1}{6}$, $q = 1 - \frac{1}{6} = \frac{5}{6}$. Воспользуемся формулой Бернулли:

$$P_{10}(2) = C_{10}^2 \cdot p^2 \cdot q^{10-2} = C_{10}^2 \cdot \left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right)^8 = \frac{10 \cdot 9}{1 \cdot 2} \cdot \frac{1}{36} \times$$
$$\times \frac{390625}{1679616} \approx 45 \cdot 0{,}028 \cdot 0{,}23 \approx 0{,}29.$$

Ответ: $\approx 0{,}29$.

Пример 2.2. Производится 6 выстрелов по цели. Вероятность попадания в цель при каждом выстреле равна 0,4. Найти вероятность того, что:
1) произойдет одно попадание в цель;
2) произойдет не менее 4 попаданий;
3) произойдет хотя бы одно попадание.

Решение
По условию, $n = 6$, $p = 0{,}4$, $q = 1 - 0{,}4 = 0{,}6$.
1. $k = 1$. По формуле Бернулли:

$$P_6(1) = C_6^1 \cdot p^1 \cdot q^{6-1} = C_6^1 \cdot 0{,}4 \cdot (0{,}6)^5 \approx 6 \cdot 0{,}4 \cdot 0{,}078 \approx$$
$$\approx 0{,}1872.$$

2. Обозначим через B событие «произойдет не менее 4 попаданий в цель». Событие B означает, что было либо четыре попадания, либо пять попаданий, либо шесть попаданий в цель. Следовательно,

$$P_6(k \ge 4) = P_6(4) + P_6(5) + P_6(6) = C_6^4 \cdot p^4 \cdot q^{6-4} + C_6^5 \cdot p^5 \times$$
$$\times q^{6-5} + C_6^6 \cdot p^6 \cdot q^{6-6} = C_6^4 \cdot (0{,}4)^4 \cdot (0{,}6)^2 + C_6^5 \cdot (0{,}4)^5 \cdot 0{,}6 +$$
$$+ C_6^6 \cdot (0{,}4)^6 \cdot (0{,}6)^0 = \frac{6 \cdot 5 \cdot 4 \cdot 3}{1 \cdot 2 \cdot 3 \cdot 4} \cdot 0{,}0256 \cdot 0{,}36 + \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} \times$$
$$\times 0{,}01024 \cdot 0{,}6 + 1 \cdot 0{,}0041 \cdot 1 \approx 15 \cdot 0{,}0092 + 6 \cdot 0{,}0061 +$$
$$+ 0{,}0041 = 0{,}138 + 0{,}0366 + 0{,}0041 = 0{,}1787.$$

3. События «произойдет хотя бы одно попадание в цель» и «из шести выстрелов нет ни одного попадания в цель» противоположны, поэтому вероятность того, что при 6 выстрелах по цели произойдет хотя бы одно попадание, равна:

$$P_6(k \ge 1) = 1 - P_6(0) = 1 - C_6^0 \cdot p^0 \cdot q^{6-0} = 1 - 1 \cdot 1 \cdot q^6 =$$
$$= 1 - q^6 = 1 - (0{,}6)^6 \approx 1 - 0{,}0467 = 0{,}9533.$$

Ответ: 1) $\approx 0{,}1872$; 2) $\approx 0{,}1787$; 3) $\approx 0{,}9533$.

Тест 2.1. Вероятность попадания в цель при одном выстреле равна 0,8. Вероятность того (по формуле Бернулли), что в серии из четырех выстрелов будет хотя бы одно попадание, равна:

1) $P_4(k \ge 1) = 1 - P_4(0)$;
2) $P_4(k \le 1) = P_4(0) + P_4(1)$;
3) $P_4(k \ge 3) = P_4(3) + P_4(4)$;
4) $P_4(0)$;
5) $P_4(1)$.

Непосредственное применение формулы Бернулли при большом числе испытаний связано с громоздкими вычислениями. Поэтому при больших n вместо нее, как правило, используют приближенные формулы Пуассона и Муавра-Лапласа.

2.2. Теорема Пуассона

Если число испытаний n достаточно велико, а вероятность p достаточно мала, причем их произведение $\lambda = np$ не мало и не велико (обычно достаточно условий $p < 0{,}1$; $npq < 10$), то вероятность $P_n(k)$ можно приближенно найти по *формуле Пуассона*:

$$P_n(k) \approx \frac{\lambda^k \cdot e^{-\lambda}}{k!}.$$

Пример 2.3. Завод отправил на базу 5000 доброкачественных изделий. Вероятность того, что в пути изделие повредится, равна 0,0002. Какова вероятность того, что на базу прибудут 3 негодных изделия?

Решение

По условию, $n = 5000$, $k = 3$, $p = 0{,}0002$. Поскольку n достаточно велико, p мала (условие $npq = 5000 \cdot 0{,}0002 \cdot (1 - 0{,}0002) = 0{,}9998 < 10$ выполнено) и $\lambda = np = 5000 \cdot 0{,}0002 = 1$, то для вычисления $P_{5000}(3)$ используем формулу Пуассона:

$$P_{5000}(3) \approx \frac{1^3 \cdot e^{-1}}{3!} = \frac{1 \cdot \frac{1}{e}}{1 \cdot 2 \cdot 3} = \frac{1}{6 \cdot e} \approx 0{,}0617.$$

Ответ: $\approx 0{,}0617$.

Тест 2.2. На факультете обучаются 1825 студентов. Для нахождения вероятности того, что 1 сентября является днем рождения одновременно четырех студентов факультета, необходимо использовать:
1) формулу Байеса;
2) формулу Пуассона;
3) классическое определение вероятности;
4) формулу полной вероятности.

2.3. Локальная теорема Муавра-Лапласа

Если число испытаний n достаточно велико, а вероятности p и q не очень близки к нулю (обычно достаточно условий $n > 100$, $npq > 20$), то вероятность $P_n(k)$ того, что в n испытаниях событие произойдет ровно k раз, можно приближенно найти по *локальной формуле Муавра-Лапласа:*

$$P_n(k) \approx \frac{1}{\sqrt{npq}} \cdot \varphi(x),$$

где $x = \dfrac{k - np}{\sqrt{npq}}$, $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$ – функция Гаусса.

Функцию $\varphi(x)$ часто называют *малой функцией Лапласа*, значения которой приведены в таблице (приложение 1).

Пользуясь таблицей, необходимо использовать *свойства функции* $\varphi(x)$:

1. Функция $\varphi(x)$ является четной, т. е. $\varphi(-x) = \varphi(x)$.

2. Функция $\varphi(x)$ – монотонно убывающая при положительных значениях x, причем при $x \to \infty$, $\varphi(x) \to 0$.

Практически можно считать, что уже при $x > 5$, $\varphi(x) \approx 0$.

Пример 2.4. Вероятность того, что изделие, сошедшее с конвейера, является изделием первого сорта, равна 0,9. Какова вероятность того, что из 400 сошедших с конвейера деталей 356 окажутся первого сорта?

Решение

По условию, $n = 400$, $k = 356$, $p = 0{,}9$, $q = 0{,}1$. Так как n достаточно велико, $npq = 400 \cdot 0{,}9 \cdot 0{,}1 = 36 > 20$, то по формуле Муавра-Лапласа:

$$P_{400}(356) \approx \frac{1}{\sqrt{400 \cdot 0{,}9 \cdot 0{,}1}} \cdot \varphi(x) = \frac{1}{\sqrt{36}} \cdot \varphi(x) = \frac{1}{6} \cdot \varphi(x).$$

Вычислим определяемое данными задачи значение x:

$$x = \frac{k - np}{\sqrt{npq}} = \frac{356 - 400 \cdot 0{,}9}{\sqrt{400 \cdot 0{,}9 \cdot 0{,}1}} = \frac{356 - 360}{\sqrt{36}} = \frac{-4}{6} \approx -0{,}67.$$

Учитывая, что $\varphi(-x) = \varphi(x)$, $\varphi(-0{,}67) = \varphi(0{,}67)$, по таблице приложения 1 находим $\varphi(0{,}67) = 0{,}3187$.

Искомая вероятность $P_{400}(356) \approx \dfrac{0{,}3187}{6} \approx 0{,}0531$.

Ответ: $\approx 0{,}0531$.

Тест 2.3. Вероятность поражения мишени при одном выстреле равна 0,8. Найти вероятность того, что при 100 выстрелах мишень будет поражена ровно 75 раз. Применяя локальную формулу Муавра-Лапласа $P_n(k) \approx \dfrac{1}{\sqrt{npq}} \cdot \varphi(x)$, находим x следующим образом:

1) $\dfrac{75 - 100 \cdot 0{,}8}{\sqrt{0{,}8 \cdot 0{,}2}}$;

2) $\dfrac{75}{\sqrt{100 \cdot 0{,}8 \cdot 0{,}2}}$;

3) $\dfrac{75 - 100 \cdot 0{,}8}{\sqrt{100 \cdot 0{,}8 \cdot 0{,}2}}$;

4) $\dfrac{75 - 0{,}8}{\sqrt{100 \cdot 0{,}8 \cdot 0{,}2}}$;

5) $\dfrac{100}{0{,}8 \cdot 75}$.

Тест 2.4. Значение функции $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$ при $x = -1{,}37$ равно:

1) $\varphi(1{,}37)$;

2) $-\varphi(1{,}37)$.

2.4. Интегральная теорема Лапласа

В условиях локальной формулы Муавра-Лапласа вероятность $P_n(k_1 \le k \le k_2)$ того, что событие появится в n испытаниях не менее k_1 и не более k_2 раз, можно приближенно найти по *интегральной формуле Муавра-Лапласа:*

$$P_n(k_1 \le k \le k_2) \approx \Phi(x_2) - \Phi(x_1),$$

где $x_1 = \dfrac{k_1 - np}{\sqrt{npq}}$, $x_1 = \dfrac{k_1 - np}{\sqrt{npq}}$, $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \cdot \displaystyle\int_0^x e^{-\frac{t^2}{2}} dt$.

Функция $\Phi(x)$ называется *стандартной,* или *нормированной, функцией Лапласа*. Таблица значений функции $\Phi(x)$ приводится в приложении 2.

При использовании этой таблицы необходимо знать *свойства функции* $\Phi(x)$:

1. Функция $\Phi(x)$ нечетная, т. е. $\Phi(-x) = -\Phi(x)$.

2. Функция $\Phi(x)$ монотонно возрастающая, причем при $x \to \infty$. $\Phi(x) \to 0{,}5$.

Практически можно считать, что уже при $x > 5$ ($x < -5$). $\Phi(x) \approx 0{,}5$, $\Phi(-x) \approx -0{,}5$.

3. $\Phi(0) = 0$.

Пример 2.5. Вероятность поражения мишени стрелком при одном выстреле равна 0,75. Найти вероятность того, что при 100 выстрелах мишень будет поражена:

1) не менее 70 и не более 80 раз;

2) не более 70 раз.

Решение

По условию, $n = 100$, $p = 0{,}75$, $q = 1 - 0{,}75 = 0{,}25$.

1. $k_1 = 70$, $k_2 = 80$. Вычислим x_1 и x_2:

$$x_1 = \frac{70-100\cdot 0{,}75}{\sqrt{100\cdot 0{,}75\cdot 0{,}25}} = \frac{70-75}{\sqrt{18{,}75}} \approx \frac{-5}{4{,}3} \approx -1{,}16; \; x_2 = \frac{80-100\cdot 0{,}75}{\sqrt{100\cdot 0{,}75\cdot 0{,}25}} = \frac{80-75}{\sqrt{18{,}75}} \approx \frac{5}{4{,}3} \approx 1{,}16.$$

Учитывая, что функция Лапласа нечетная, получим $\Phi(x_1) = \Phi(-1{,}16) = = -\Phi(1{,}16)$. По таблице приложения 2 находим: $\Phi(1{,}16)=0{,}3770$. Искомая вероятность:

$$P_{100}(70 \le k \le 80) = P_{100}(70{,}80) \approx \Phi(1{,}16) - \Phi(-1{,}16) =$$
$$= \Phi(1{,}16) + \Phi(1{,}16) = 2\cdot 0{,}3770 = 0{,}754.$$

2. Требование, чтобы мишень была поражена не более 70 раз, означает, что она может быть поражена 0 раз, либо 1 раз, …, либо 70 раз. Таким образом, в рассматриваемом случае следует применять $k_1 = 0$, $k_2 = 70$. Тогда

$$x_1 = \frac{0-100\cdot 0{,}75}{\sqrt{100\cdot 0{,}75\cdot 0{,}25}} = \frac{75}{\sqrt{18{,}75}} \approx \frac{75}{4{,}3} \approx -17{,}44;$$

$$x_2 = \frac{70-100\cdot 0{,}75}{\sqrt{100\cdot 0{,}75\cdot 0{,}25}} = \frac{70-75}{\sqrt{18{,}75}} \approx \frac{-5}{4{,}3} \approx -1{,}16.$$

Так как $x_1 \approx -17{,}44 < -5$, то $\Phi(-17{,}44) \approx -0{,}5$. Итак,

$$P_{100}(0 \le k \le 70) \approx \Phi(-1{,}16) - \Phi(-17{,}44) = -0{,}3770 + 0{,}5 = 0{,}123.$$

Ответ: 1) $\approx 0{,}754$; 2) $\approx 0{,}123$.

Тест 2.5. В каждом из 500 независимых испытаний событие A происходит с постоянной вероятностью 0,4. Вероятность того, что событие A происходит не менее 180 и не более 220 раз, по интегральной теореме Лапласа находим следующим образом:

1) $\dfrac{220-500\cdot 0{,}4}{\sqrt{500\cdot 0{,}4\cdot 0{,}6}} - \dfrac{180-500\cdot 0{,}4}{\sqrt{500\cdot 0{,}4\cdot 0{,}6}}$;

2) $\dfrac{220-500\cdot 0{,}4}{\sqrt{500\cdot 0{,}4\cdot 0{,}6}}$;

3) $\dfrac{220-0{,}4}{\sqrt{500\cdot 0{,}6}} - \dfrac{180-0{,}4}{\sqrt{500\cdot 0{,}6}}$;

4) $\dfrac{180-500\cdot 0{,}4}{\sqrt{500\cdot 0{,}4\cdot 0{,}6}} - \dfrac{220-500\cdot 0{,}4}{\sqrt{500\cdot 0{,}4\cdot 0{,}6}}$;

5) $C_{220}^{180}\cdot 0{,}4^{180}\cdot 0{,}6^{220-180}$.

Тест 2.6. Значение функции Лапласа

$\Phi(x)=\frac{1}{\sqrt{2\pi}}\cdot\int_0^x e^{-\frac{t^2}{2}}dt$ при $x = -6$ равно:

1) $\Phi(6)$;

2) $-\Phi(6)$.

2.5. Наивероятнейшее число появлений события в независимых испытаниях

Число k_0 (наступления события в независимых испытаниях, в каждом из которых вероятность появления события равна p) называется *наивероятнейшим,* если вероятность того, что событие наступит в этих испытаниях k_0 раз, превышает (или, по крайней мере, не меньше) вероятности остальных возможных исходов испытаний.

Наивероятнейшее число k_0 определяют из двойного неравенства

$$np-q\le k_0\le np+p,$$

причем:

1) если число $np - q$ дробное, то существует одно наивероятнейшее число k_0;

2) если число $np - q$ целое, то существует два наивероятнейших числа, а именно: k_0 и $k_0 + 1$;

3) если число np целое, то наивероятнейшее число $k_0 = np$.

Пример 2.6. Доля изделий высшего сорта на предприятии составляет 40%. Чему равно наивероятнейшее число изделий высшего сорта в случайно отобранной партии из 120 изделий?

Решение

По условию $n = 120$, $p=\frac{40}{100}=0{,}4$, $q = 1 - 0{,}4 = 0{,}6$.

Наивероятнейшее число k_0 находим из двойного неравенства

$$np-q\le k_0\le np+p.$$

Подставим данные задачи: $120\cdot 0{,}4 - 0{,}6 \le k_0 \le 120\cdot 0{,}4 + 0{,}4$, получаем $47{,}4\le k_0\le 48{,}4$.

Так как k_0 целое число, заключенное между 47,4 и 48,4 , то $k_0 = 48$.

Ответ: 48.

Пример 2.7. Вероятность попадания в цель при одном выстреле $p = 0{,}7$. Какова вероятность наивероятнейшего числа попаданий, если произведено 9 выстрелов?

Решение

Находим наивероятнейшее число попаданий из двойного неравенства: $9\cdot 0{,}7 - 0{,}3 \le k_0 \le 9\cdot 0{,}7 - 0{,}7$, получаем $6 \le k_0 \le 7$.

Получили два наивероятнейших числа: $k_0 = 6$ и $k_0 = 7$. Вероятности их наибольшие и равны между собой. Найдем одно из этих значений, например,

$$P_9(6)=C_9^6\cdot p^6\cdot q^{9-6}=C_9^6\cdot(0{,}7)^6\cdot(0{,}3)^3\approx\frac{9\cdot 8\cdot 7\cdot 6\cdot 5\cdot 4}{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6}\times$$
$$\times 0{,}12\cdot 0{,}027\approx 84\cdot 0{,}003=0{,}252.$$

Ответ: $\approx 0{,}252$.

Тест 2.7. В результате многолетних наблюдений для некоторой местности было установлено, что вероятность выпадения дождя 13 июля равна $\frac{4}{17}$. Наивероятнейшее число дождливых дней 13 июля в ближайшие 50 лет равно:

1) $\frac{4}{17}\le k_0\le\frac{13}{17}$;

2) $50\cdot\frac{4}{17}-\frac{4}{17}\le k_0\le 50\cdot\frac{4}{17}+\frac{13}{17}$;

3) $50\cdot\frac{4}{17}\le k_0\le 50\cdot\frac{13}{17}$;

4) $50\cdot\frac{4}{17}\le k_0\le 50\cdot\frac{4}{17}+\frac{4}{17}$;

5) $50\cdot\frac{4}{17}-\frac{13}{17}\le k_0\le 50\cdot\frac{4}{17}+\frac{4}{17}$.

Вопросы для самоконтроля

1. Какой вид имеет формула Бернулли?
2. Когда применяются локальная и интегральная теоремы Лапласа?
3. В каких случаях нужно пользоваться формулой Пуассона?
4. Как определяется малая функция Лапласа? Каковы ее свойства?
5. Как определяется функция Лапласа? Каковы свойства функции Лапласа?
6. Что называется наивероятнейшим числом наступления события в n независимых испытаниях?

Ответы на тестовые задания

Номер теста	2.1	2.2	2.3	2.4	2.5	2.6	2.7
Правильный ответ	1	2	3	1	1	2	5

3. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ, ИХ РАСПРЕДЕЛЕНИЕ И ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ

3.1. Понятие случайной величины. Классификация случайных величин

Случайной величиной называют величину X, которая в результате испытания принимает одно значение из множества возможных значений, заранее неизвестное и зависящее от случая.

Случайные величины обозначаются заглавными буквами латинского алфавита X, Y, Z, …

Пример 3.1. Число студентов на лекции – случайная величина.

Дискретной случайной величиной называется случайная величина, принимающая изолированное значение из некоторого конечного или бесконечного промежутка. Число ее значений счетно.

Пример 3.2. Случайная величина X – число студентов на лекции. Данная случайная величина является дискретной.

Непрерывной случайной величиной называется случайная величина, которая может принимать все значения из некоторого конечного или бесконечного промежутка.

Пример 3.3. Случайная величина X – время финиша в соревнованиях. Данная случайная величина является непрерывной.

Тест 3.1. Пусть случайная величина X – число солнечных дней в месяце. Она является:
1) дискретной случайной величиной;
2) непрерывной случайной величиной.

Одномерной случайной величиной называется случайная величина, каждое значение которой является числом. Она изображается точкой на числовой прямой.

Пример 3.4. Случайная величина X – число студентов на лекции. Данная случайная величина является одномерной дискретной случайной величиной.

Пример 3.5. Случайная величина X – время финиша в соревнованиях. Данная случайная величина является одномерной непрерывной случайной величиной.

Количественной случайной величиной называется случайная величина, значение которой может быть измерено по какой-либо шкале.

Пример 3.6. Случайная величина X – оценка на экзамене. Данная случайная величина является количественной случайной величиной.

Качественной случайной величиной называется случайная величина, значение которой нельзя измерить по какой-либо шкале.

Пример 3.7. Случайная величина X – знания студентов. Данная случайная величина является качественной случайной величиной.

Любая качественная случайная величина методом экспертных оценок может быть преобразована в количественную, например, знания оцениваются баллами.

Любая случайная величина задается законом распределения.

3.2. Закон распределения дискретной случайной величины

Законом распределения дискретной случайной величины называется всякое соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями.

Закон распределения дискретной случайной величины может быть задан в виде ряда распределения, аналитически, графически, функцией распределения.

Пусть X – дискретная случайная величина, x_1; x_2; …; x_n – ее значения. Каждому x_i поставим в соответствие ее вероятность, т. е. $P(X = x_i) = p_i$, где $i = 1; 2;…; n$ (аналитический вид закона распределения).

Рядом распределения дискретной случайной величины называется таблица

x_i	x_1	x_2	…	x_n
p_i	p_1	p_2	…	p_n

в которой перечислены возможные значения x_1; x_2; …; x_n случайной величины и соответствующие им вероятности p_1; p_2; …; p_n. Так как события «$X = x_1$»; «$X = x_2$»; …; «$X = x_n$» образуют полную группу, то $\sum_{i=1}^{n} p_i = 1$.

Графически случайная величина задается в виде полигона или многоугольника распределения.

Полигоном распределения вероятностей называется ломаная линия, состоящая из отрезков, соединяющих точки с координатами $(x_i; p_i)$.

Многоугольником распределения называется часть плоскости, ограниченная полигоном распределения вероятностей и прямыми $x = x_1$, $x = x_n$ и $y = 0$.

Функцией распределения случайной величины X называется вероятность $P(X < x)$ того, что случайная величина X примет значение, меньшее x, т. е. $F(X) = P(X < x)$.

Значение функции распределения дискретной случайной величины X вычисляется по формуле

$$F(x) = \sum_{x_i < x} P(X = x_i),$$

где суммирование ведется по всем значениям i, для которых $x_i < x$.

Основные свойства функции распределения:

1) $0 \le F(x) \le 1$;

2) функция распределения монотонно возрастает, т. е. если $x_1 < x_2$, то $F(x_1) \le F(x_2)$;

3) функция распределения непрерывна слева;

4) $\lim\limits_{x \to -\infty} F(x) = 0$; $\lim\limits_{x \to +\infty} F(x) = 1$.

Пример 3.8. Монету бросают 4 раза. Случайная величина X – число появлений герба. Требуется выполнить следующее:

1) построить ряд распределения;

2) построить полигон и многоугольник распределения;

3) найти функцию распределения и построить ее график.

Решение

1. Так как монету подбрасывают 4 раза, то герб может появится либо все 4 раза, либо 3 раза, либо 2 раза, либо 1 раз, либо не появится, т. е. 4 раза выпадет цифра. Поэтому возможные значения случайной величины X : $x_1 = 0$; $x_2 = 1$; $x_3 = 2$; $x_4 = 3$; $x_5 = 4$. Поскольку подбрасывания монеты 4 раза являются повторными независимыми испытаниями относительно появления герба, то вероятность возможных значений случайной величины X находится по формуле Бернулли

$$P_n(k) = C_n^k \cdot p^k \cdot q^{n-k},$$

где $n = 4$ – число всех испытаний;

$k = 0;1;2;3;4$ – число возможных появлений герба;

$p = \dfrac{1}{2}$ – вероятность появления герба в одном испытании, т. е. при одном бросании монеты;

$q = 1 - p$; $q = 1 - \dfrac{1}{2} = \dfrac{1}{2}$ – вероятность противоположного события, т. е. выпадение цифры в одном испытании.

Если $x_1 = 0$, то

$$p_0 = P(X = 0) = P_4(0) = C_4^0 \cdot \left(\frac{1}{2}\right)^0 \cdot \left(\frac{1}{2}\right)^{4-0} = 1 \cdot 1 \cdot \left(\frac{1}{2}\right)^4 = \frac{1}{16}.$$

Если $x_2 = 1$, то

$$p_1 = P(X = 1) = P_4(1) = C_4^1 \cdot \left(\frac{1}{2}\right)^1 \cdot \left(\frac{1}{2}\right)^{4-1} = \frac{4!}{1!(4-1)!} \cdot \left(\frac{1}{2}\right)^1 \cdot \left(\frac{1}{2}\right)^3 = = 4 \cdot \left(\frac{1}{2}\right)^4 = \frac{4}{16}.$$

Если $x_3 = 2$, то

$$p_2 = P(X = 2) = P_4(2) = C_4^2 \cdot \left(\frac{1}{2}\right)^2 \cdot \left(\frac{1}{2}\right)^{4-2} = \frac{4!}{2!(4-2)!} \cdot \left(\frac{1}{2}\right)^2 \cdot \left(\frac{1}{2}\right)^2 =$$

$$= \frac{3 \cdot 4}{1 \cdot 2} \cdot \left(\frac{1}{2}\right)^4 = \frac{6}{16}.$$

Если $x_4 = 3$, то

$$p_3 = P(X = 3) = P_4(3) = C_4^3 \cdot \left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^{4-3} = \frac{4!}{3!(4-3)!} \cdot \left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^1 =$$

$$= \frac{4}{1} \cdot \left(\frac{1}{2}\right)^4 = \frac{4}{16}.$$

Если $x_5 = 4$, то

$$p_4 = P(X = 4) = P_4(4) = C_4^4 \cdot \left(\frac{1}{2}\right)^4 \cdot \left(\frac{1}{2}\right)^{4-4} = 1 \cdot \left(\frac{1}{2}\right)^4 \cdot \left(\frac{1}{2}\right)^0 = \frac{1}{16}.$$

Контроль вычислений: $\sum\limits_{i=0}^{4} p_i = \dfrac{1}{16} + \dfrac{4}{16} + \dfrac{6}{16} + \dfrac{4}{16} + \dfrac{1}{16} = \dfrac{16}{16} = 1$.

Таким образом, ряд распределения случайной величины X имеет вид

X	0	1	2	3	4
p_i	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

2. В прямоугольной системе координат строим точки с координатами $\left(0; \dfrac{1}{16}\right)$, $\left(1; \dfrac{4}{16}\right)$, $\left(2; \dfrac{6}{16}\right)$; $\left(3; \dfrac{4}{16}\right)$, $\left(4; \dfrac{1}{16}\right)$, соединяем эти точки отрезками.

Полученная ломаная является полигоном распределения случайной величины X (рис. 3).

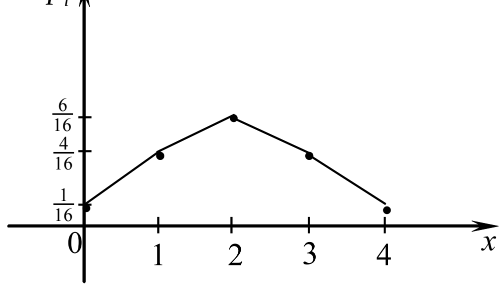

Рис. 3

Многоугольник распределения приведен на рис. 4.

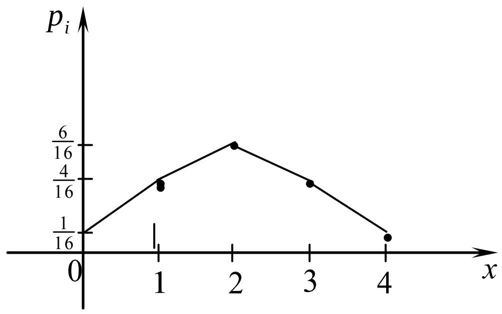

Рис. 4

3. Найдем функцию распределения случайной величины X.

Функция $F(x)$ определена для всех $x \in (-\infty; +\infty)$. Значения случайной величины $x_1 = 0$, $x_2 = 1$, $x_3 = 2$, $x_4 = 3$, $x_5 = 4$ разбивают числовую прямую на 6 интервалов (рис. 5).

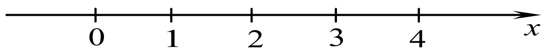

Рис. 5

Для значений x, принадлежащих интервалу $(x_j; x_i)$, $F(x) = \sum\limits_{x_i < x} P(X < x_i)$. Найдем значения $F(x)$ на каждом интервале.

Пусть $x \in (-\infty; 0]$. В интервал $(-\infty; x)$ не попадает ни одно значение случайной величины X (рис. 6).

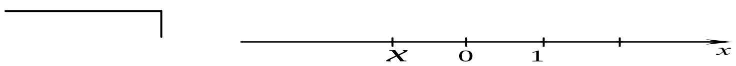

Рис. 6

Значит, $F(x) = \sum\limits_{x_i < x} P(X < x_i) = 0$.

Пусть $x \in (0; 1]$. Условию $X < x$ при $x \in (0; 1]$ удовлетворяет только одно значение $X = 0$ (рис. 7) с вероятностью $\frac{1}{16}$, поэтому $F(x) = \sum\limits_{x_i < x} P(X = x_i) = 0 + \frac{1}{16} = \frac{1}{16}$.

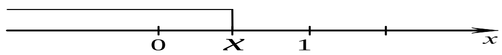

Рис. 7

Пусть $x \in (1; 2]$. Условию $X < x$ при $x \in (1; 2]$ удовлетворяют два значения $X = 0$, $X = 1$ с вероятностями $\frac{1}{16}$ и $\frac{4}{16}$ соответственно, поэтому $F(x) = \frac{1}{16} + \frac{4}{16} = \frac{5}{16}$.

Аналогично, если $x \in (2; 3]$, то $F(x) = \frac{1}{16} + \frac{4}{16} + \frac{6}{16} = \frac{11}{16}$.

Если $x \in (3; 4]$, то $F(x) = \frac{1}{16} + \frac{4}{16} + \frac{6}{16} + \frac{4}{16} = \frac{15}{16}$.

Если $x \in (4; +\infty)$, то $F(x) = \frac{1}{16} + \frac{4}{16} + \frac{6}{16} + \frac{4}{16} + \frac{1}{16} = 1$.

Таким образом,

$$F(x)=\begin{cases} 0, & \text{если } x \le 0, \\ \dfrac{1}{16}, & \text{если } 0 < x \le 1, \\ \dfrac{5}{16}, & \text{если } 1 < x \le 2, \\ \dfrac{11}{16}, & \text{если } 2 < x \le 3, \\ \dfrac{15}{16}, & \text{если } 3 < x \le 4, \\ 1, & \text{если } x > 4. \end{cases}$$

График этой функции представлен на рис. 8.

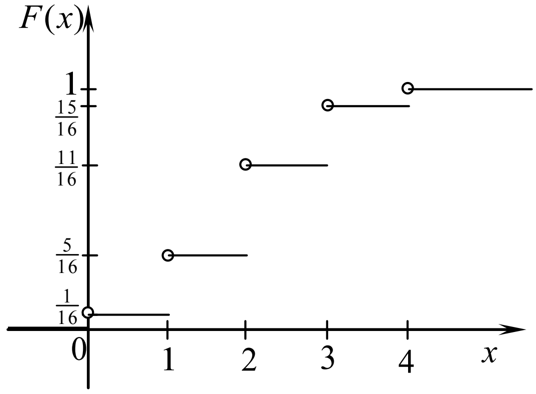

Рис. 8

Тест 3.2. Случайная величина задана рядом распределения

X	-2	1	5
p_i	?	0,3	0,1

Вместо знака «?» следует поставить число:
1) 0;
2) 0,6;
3) 0,2;
4) 0,1;
5) 0,3.

Тест 3.3. В партии из 10 деталей 8 стандартных. Наудачу отобраны 2 детали. Случайная величина X – число стандартных деталей среди отобранных. Все возможные значения случайной величины X включают:
1) {0};
2) {0;1};
3) {1;2};
4) {0;1;2};
5) {0;1;2;3;4;5;6;7;8}.

Тест 3.4. Случайная величина, задана рядом распределения

X	-2	1	5
p_i	0,6	0,3	0.1

Если x $\in$ [1;5], $F(x)$ равно:
1) 0;
2) 0,6;
3) 0,3;
4) 1;
5) 0,9.

3.3. Числовые характеристики дискретных случайных величин

Математическим ожиданием дискретной случайной величины называется сумма произведений всех ее возможных значений на их вероятности:

$$M(X) = x_1 \cdot p_1 + x_2 \cdot p_2 + ... + x_n \cdot p_n. \qquad (3.1)$$

Математическое ожидание оценивает среднее значение случайной величины.

Свойства математического ожидания:

1. $M(C) = C$.
2. $M(CX) = CM(X)$.
3. $M(X + Y) = M(X) + M(Y)$.
4. $M(XY) = M(X)M(Y)$, случайные величины X и Y независимые.

Отклонением случайной величины X от ее математического ожидания M(X) называется случайная величина $X - M(X)$.

Дисперсией случайной величины X называется математическое ожидание квадрата отклонения случайной величины от ее математического ожидания:

$$D(X) = M\left(X - M(X)\right)^2. \qquad (3.2)$$

Дисперсия характеризует степень разброса значений случайной величины относительно математического ожидания.

Свойства дисперсии:

1. $D(C) = 0$.
2. $D(CX) = C^2D(X)$.
3. $D(X \pm Y) = D(X) + D(Y)$.

Дисперсию случайной величины удобно вычислять по формуле

$$D(X) = M(X^2) - \left(M(X)\right)^2. \qquad (3.3)$$

Средним квадратическим отклонением случайной величины называется корень квадратный из дисперсии, т. е.

$$\sigma(X) = \sqrt{D(X)}. \qquad (3.4)$$

Так же как и дисперсия среднее квадратическое отклонение характеризует степень разброса значений случайной величины относительно математического ожидания. Единицы измерения $M(X)$ и $\sigma(X)$ совпадают, $D(X)$ измеряется в единицах квадратных.

Пример 3.9. Найти математическое ожидание, дисперсию, среднее квадратическое отклонение дискретной случайной величины, ряд распределения которой получили в примере 3.8.

Решение

В примере 3.8 получили приведенный ниже ряд распределения

X	0	1	2	3	4
p_i	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

1. По формуле (3.1) найдем математическое ожидание:

$$M(X) = 0 \cdot \frac{1}{16} + 1 \cdot \frac{4}{16} + 2 \cdot \frac{6}{16} + 3 \cdot \frac{4}{16} + 4 \cdot \frac{1}{16} = 0 + \frac{4}{16} + \frac{12}{16} +$$
$$+ \frac{12}{16} + \frac{4}{16} = \frac{32}{16} = 2.$$

2. По формуле (3.3) определим дисперсию:

$$D(X) = M(X^2) - \left(M(X)\right)^2.$$
$$D(X) = 0^2 \cdot \frac{1}{16} + 1^2 \cdot \frac{4}{16} + 2^2 \cdot \frac{6}{16} + 3^2 \cdot \frac{4}{16} + 4^2 \cdot \frac{1}{16} - 2^2 = 1.$$

3. По формуле (3.4) найдем среднее квадратическое отклонение:

$$\sigma(X) = \sqrt{1} = 1.$$

Пример 3.10. Найти математическое ожидание случайной величины $Y = 3X + 5$, если $M(X) = 2$.

Решение
Используя свойства математического ожидания, получим:

$$M(Y) = M(3X + 5) = M(3X) + M(5) = 3M(X) + 5 = 3 \cdot 2 + 5 = 11$$

Ответ: 11.

Пример 3.11. Найти дисперсию случайной величины $Y = 3X + 5$, если $D(X) = 4$.

Решение
Используя свойства дисперсии, получим:

$$D(Y) = D(3X + 5) = D(3X) + D(5) = 3^2 D(X) + 0 = 9 \cdot 4 = 36 .$$

Ответ: 36.

Тест 3.5. Математическое ожидание дискретной случайной величины, заданной рядом распределения

X	-1	0	1
p_i	0,1	0,7	0,2

равно:

1) –2;

2) 0;

3) 1;

4) 0,014;

5) 0,1.

Тест 3.6. Дисперсия дискретной случайной величины, заданной рядом распределения

X	-1	0	1
p_i	0,1	0,7	0,2

равна:

1) $D(X) = (-1)^2 \cdot 0{,}1 + 0^2 \cdot 0{,}7 + 1^2 \cdot 0{,}2$;

2) $D(X) = (-1) \cdot 0{,}1 + 0 \cdot 0{,}7 + 1 \cdot 0{,}2$;

3) $D(X) = (-1)^2 \cdot 0{,}1 + 0^2 \cdot 0{,}7 + 1^2 \cdot 0{,}2 - (-1) - 0 - 1$;

4) $D(X) = (-1)^2 \cdot 0{,}1 + 0^2 \cdot 0{,}7 + 1^2 \cdot 0{,}2 - ((-1) \cdot 0{,}1 + 0 \cdot 0{,}7 + 1 \cdot 0{,}2)^2$;

5) $D(X) = (-1 \cdot 0{,}1 + 0 \cdot 0{,}7 + 1 \cdot 0{,}2) - (-1)^2 \cdot 0{,}1 - 0^2 \cdot 0{,}7 - 1^2 \cdot 0{,}2$.

Тест 3.7. Известно математическое ожидание случайной величины X : $M(X) = 1$. Тогда математическое ожидание случайной величины $Y = 2 + 7X$ будет равно:

1) 2;
2) 9;
3) 1;
4) 7;
5) 0.

Тест 3.8. Дисперсия случайной величины X равна 2. Тогда дисперсия случайной величины $Y = 5X - 1$ будет равна:

1) 2;
2) 9;
3) 49;
4) 10;
5) 50.

3.4. Непрерывные случайные величины

Функция распределения непрерывной случайной величины $F(x) = P(X < x)$ является непрерывно дифференцируемой, за исключением конечного числа точек.

Все свойства функции распределения дискретной случайной величины выполняются и для функции распределения непрерывной случайной величины.

Производная от функции распределения $F(x)$ называется *плотностью распределения вероятностей* (или дифференциальной функцией распределения):

$$f(x) = F'(x).$$

Пример 3.12. Дана функция распределения непрерывной случайной величины

$$F(x) = \begin{cases} 0, & при\ x \le 5, \\ \dfrac{x}{5} - 1, & при\ 5 < x \le 10, \\ 1, & при\ x > 10. \end{cases}$$

Найти плотность распределения этой случайной величины.

Решение

$$f(x) = F'(x) = \begin{cases} 0, & при\ x < 5, \\ \dfrac{1}{5}, & при\ 5 < x < 10, \\ 0, & при\ x > 10. \end{cases}$$

При $x = 5$ имеем: $\lim\limits_{x\to 5-0} \dfrac{\Delta F}{\Delta x} = 0$, $\lim\limits_{x\to 5+0} \dfrac{\Delta F}{\Delta x} = \dfrac{1}{5}$.

Так как $0 \ne \dfrac{1}{5}$, то $F'(5)$ не существует.

При $x = 10$ имеем: $\lim\limits_{x\to 10-0} \dfrac{\Delta F}{\Delta x} = \dfrac{1}{5}$, $\lim\limits_{x\to 10+0} \dfrac{\Delta F}{\Delta x} = 0$.

Так как $\dfrac{1}{5} \ne 0$, то $F'(10)$ не существует.

Тест 3.9. Дана функция распределения непрерывной случайной величины X:

$$F(x) = \begin{cases} 0,\ при\ x \le 0; \\ \sin 2x;\ при\ 0 < x \le \dfrac{\pi}{4}; \\ 1;\ при\ x > \dfrac{\pi}{4}. \end{cases}$$

Дифференциальная функция распределения (плотность распределения) $f(x)$ будет равна:

1) 0;

2) 1;

3) $f(x) = \begin{cases} 0,\ при\ x < 0; \\ \cos 2x;\ при\ 0 < x \le \dfrac{\pi}{4}; \\ 0;\ при\ x > \dfrac{\pi}{4}. \end{cases}$

4) $f(x) = \begin{cases} 0,\ при\ x < 0; \\ 2\cos 2x;\ при\ 0 < x \le \dfrac{\pi}{4}; \\ 0;\ при\ x > \dfrac{\pi}{4}. \end{cases}$

5) $f(x)=\begin{cases} 0, \, при \, x<0; \\ 2\cos 2x; \, при \, 0<x\le\dfrac{\pi}{4}; \\ 1; \, при \, x>\dfrac{\pi}{4}. \end{cases}$

Зная плотность распределения, можно найти функцию распределения по формуле

$$F(x)=\int_{-\infty}^{x} f(x)\,dx.$$

Вероятность того, что непрерывная случайная величина X примет значение, принадлежащее интервалу $(a;b)$, определяется равенством

$$P(a<X<b)=\int_a^b f(x)\,dx \text{ через } f(x),$$

$$P(a<X<b)=F(b)-F(a) \text{ через } F(x).$$

Пример 3.13. Дана функция распределения непрерывной случайной величины X

$$F(x)=\begin{cases} 0, \; при \; x\le 0, \\ \sin 2x, \; при \; 0<x\le\dfrac{\pi}{4}, \\ 1, \; при \; x>\dfrac{\pi}{4}. \end{cases}$$

Найти вероятность того, что X примет значение, принадлежащее интервалу $\left(0;\dfrac{\pi}{6}\right)$, через $F(x)$ и $f(x)$.

Решение

1. Воспользуемся формулой

$$F(a<X<b)=F(b)-F(a).$$

По условию $a=0$; $b=\dfrac{\pi}{6}$; на этом интервале $F(x)=\sin 2x$. Следовательно, искомая вероятность

$$P\left(0<X<\frac{\pi}{6}\right)=F\left(\frac{\pi}{6}\right)-F(0)=\sin\left(2\cdot\frac{\pi}{6}\right)-\sin(2\cdot 0)=$$

$$=\sin\frac{\pi}{3}-\sin 0=\frac{\sqrt{3}}{2}-0=\frac{\sqrt{3}}{2}.$$

2. Найдем плотность распределения:

$$f(x)=F'(x)=\begin{cases} 0, \; при \; x<0, \\ 2\cdot\cos 2x, \; при \; 0<x\le\dfrac{\pi}{4}, \\ 0, \; при \; x>\dfrac{\pi}{4}. \end{cases}$$

Воспользуемся формулой $P(a<X<b)=\int_a^b f(x)\,dx$.

Тогда $F\left(0<X<\dfrac{\pi}{6}\right)=\int_0^{\frac{\pi}{6}} 2\cdot\cos 2x\,dx=2\cdot\int_0^{\frac{\pi}{6}}\cos 2x\,dx=$

$=2\cdot\dfrac{1}{2}\cdot\sin 2x\Big|_0^{\frac{\pi}{6}}=\sin 2x\Big|_0^{\frac{\pi}{6}}=\sin\left(2\cdot\dfrac{\pi}{6}\right)-\sin(2\cdot 0)=\sin\dfrac{\pi}{3}-\sin 0=$

$=\dfrac{\sqrt{3}}{2}-0=\dfrac{\sqrt{3}}{2}.$

Ответ: $\dfrac{\sqrt{3}}{2}$.

Тест 3.10. Дана функция распределения непрерывной случайной величины

$$F(x)=\begin{cases} 0, & при\ x\le 0,\\ \dfrac{x}{5}, & при\ 0<x\le 5,\\ 1, & при\ x>5.\end{cases}$$

Вероятность того, что случайная величина примет значение, заключенное в интервале (3;4), равна:
1) 0;
2) $\frac{1}{5}$;
3) $\frac{2}{5}$
4) 1;
5) $\frac{4}{5}$.

3.5. Числовые характеристики непрерывной случайной величины

Математическое ожидание непрерывной случайной величины X, возможные значения которой принадлежат всей оси *Ox*, определяется равенством

$$M(X)=\int\limits_{-\infty}^{+\infty} x\cdot f(x)dx,$$

где $f(x)$ – плотность распределения.

В частности, если все возможные значения принадлежат интервалу $(a;b)$, то $M(X)=\int\limits_a^b x\,f(x)dx$.

Дисперсия непрерывной случайной величины X, возможные значения которой принадлежат оси *Ox*, определяется равенством

$$D(X)=\int\limits_{-\infty}^{+\infty}(x-M(X))^2 f(x)dx$$

или равносильным равенством

$$D(X)=\int\limits_{-\infty}^{+\infty}x^2\cdot f(x)dx-(M(X))^2 .$$

В частности, если все возможные значения принадлежат интервалу $(a;b)$, то

$$D(X)=\int\limits_a^b(x-M(X))^2\cdot f(x)dx$$

или

$$D(X)=\int\limits_a^b x^2\cdot f(x)dx-(M(X))^2 .$$

Среднее квадратическое отклонение непрерывной случайной величины определяется так же, как и для дискретной случайной величины

$$\sigma(X)=\sqrt{D(X)} .$$

Пример 3.14. Непрерывная случайная величина задана функцией распределения

$$F(x)=\begin{cases} 0,\ при\ \ x\le 0,\\ \dfrac{1}{2}-\dfrac{1}{2}\cdot\cos x, при\ \ 0<x\le\pi,\\ 1,\ при\ \ x>\pi.\end{cases}$$

Найти математическое ожидание, дисперсию, среднее квадратическое отклонений.

Решение
1. Найдем плотность распределения:

$$f(x)=F'(x)=\begin{cases}0, \text{ при } x\le 0,\\ \dfrac{1}{2}\cdot\sin x, \text{ при } 0<x\le\pi,\\ 0, \text{ при } x>\pi.\end{cases}$$

2. Для вычисления математического ожидания воспользуемся формулой

$$M(X)=\int_a^b x\cdot f(x)\,dx,$$

где a и b – концы интервала, в котором заключены возможные значения X.

По условию $a=0$; $b=\pi$; $f(x)=\dfrac{1}{2}\sin x$.

Следовательно, имеем:

$$M(X)=\int_0^\pi x\cdot\frac{1}{2}\sin x\,dx=\frac{1}{2}\cdot\int_0^\pi x\cdot\sin x\,dx==\left[\begin{matrix}\textit{воспользуемся методом}\\ \textit{интегрирования по частям}\\ u=x;\quad dv=\sin x\,dx\\ du=dx;\ v=\int\sin x\,dx=-\cos x\end{matrix}\right]=$$

$$=\frac{1}{2}\cdot\left(\left(x(-\cos x)\right)\Big|_0^\pi-\int_0^\pi(-\cos x)\,dx\right)=\frac{1}{2}\cdot\left(\left(-x\cdot\cos x\right)\Big|_0^\pi+\int_0^\pi\cos x\,dx\right)=$$

$$=\frac{1}{2}\cdot\left(\left(-x\cdot\cos x\right)\Big|_0^\pi+\left(\sin x\right)\Big|_0^\pi\right)=$$

$$=\frac{1}{2}\cdot\left(\left(-\pi\cdot\cos\pi-(-0\cdot\cos 0)\right)+\left(\sin\pi-\sin 0\right)\right)=$$

$$=\frac{1}{2}\cdot\left(\left(-\pi\cdot(-1)-0\right)+\left(0-0\right)\right)=\frac{1}{2}\cdot\pi.$$

3. Дисперсию вычисляем по формуле

$$D(X)=\int_a^b x^2\cdot f(x)\,dx-\left(M(X)\right)^2.$$

$$D(X)=\int_0^\pi x^2\cdot\frac{1}{2}\cdot\sin x\,dx-\left(\frac{1}{2}\cdot\pi\right)^2=\frac{1}{2}\cdot\int_0^\pi x^2\cdot\sin x\,dx-\frac{1}{4}\cdot\pi^2=$$

$$=\left[\begin{matrix}u=x^2; & dv=\sin x\,dx\\ du=2x\cdot dx; & v=\int\sin x\,dx=-\cos x\end{matrix}\right]=$$

$$=\frac{1}{2}\cdot\left(\left(x^2\cdot(-\cos x)\right)\Big|_0^\pi-\int_0^\pi(-\cos x)\cdot 2x\,dx\right)-\frac{1}{4}\cdot\pi^2=$$

$$=\frac{1}{2}\cdot\left(\left(-x^2\cdot\cos x\right)\Big|_0^\pi+2\cdot\int_0^\pi x\cdot\cos x\,dx\right)-\frac{1}{4}\cdot\pi^2=$$

$$=\left[\begin{matrix}u=x; & dv=\cos x\,dx\\ du=dx; & v=\sin x\end{matrix}\right]=$$

$$=\frac{1}{2}\cdot\left(\left(-x^2\cdot\cos x\right)\Big|_0^\pi+2\cdot\left(\left(x\cdot\sin x\right)\Big|_0^\pi-\int_0^\pi\sin x\,dx\right)\right)-\frac{1}{4}\cdot\pi^2=$$

$$=\frac{1}{2}\cdot\left(-x^2\cdot\cos x\Big|_0^\pi+2\cdot\left(\left(x\cdot\sin x\right)\Big|_0^\pi+\cos x\Big|_0^\pi\right)\right)-\frac{1}{4}\cdot\pi^2=$$

$$=\frac{1}{2}\cdot\left(-x^2\cdot\cos x+2x\cdot\sin x+2\cdot\cos x\right)\Big|_0^\pi-\frac{1}{4}\cdot\pi^2=\frac{1}{2}\cdot\left((-\pi^2\cdot\cos\pi+\right.$$

$$\left.+2\cdot\pi\cdot\sin\pi+2\cdot\cos\pi)-(-0^2\cdot\cos 0+2\cdot 0\cdot\sin 0+2\cdot\cos 0)\right)-\frac{1}{4}\cdot\pi^2=$$

$$=\frac{1}{2}\cdot\left(\left(-\pi^2\cdot(-1)+2\cdot\pi\cdot 0+2\cdot(-1)\right)-(0+0+2)\right)-\frac{1}{4}\cdot\pi^2=$$

$$=\frac{1}{2}\cdot\left(\pi^2-2-2\right)-\frac{1}{4}\cdot\pi^2=\frac{1}{2}\cdot\pi^2-2-\frac{1}{4}\cdot\pi^2=\frac{1}{4}\cdot\pi^2-2 .$$

4. Вычисляем среднее квадратическое отклонение.

$$\sigma(X)=\sqrt{\frac{1}{4}\cdot\pi^2-2} .$$

Тест 3.11. Непрерывная случайная величина задана плотностью распределения

$$f(x)=\begin{cases}0, \textit{при}\, x<0;\\ 2\cos 2x;\textit{при}\, 0<x\le\frac{\pi}{4};\\ 0;\textit{при}\, x>\frac{\pi}{4}.\end{cases}$$

Математическое ожидание и дисперсия случайной величины равны:

1) $M(X)=\int\limits_0^{\frac{\pi}{4}} x\cdot 2\cos 2x dx$, $D(X)=\int\limits_0^{\frac{\pi}{4}} x^2 2\cos 2x dx-\left(M(x)\right)^2$;

2) $M(X)=\int\limits_0^{\frac{\pi}{4}} x\cdot 2\cos 2x dx$, $D(X)=\int\limits_0^{\frac{\pi}{4}} x^2 2\cos 2x dx-\left(M(x)\right)$;

3) $M(X)=\int\limits_0^{\frac{\pi}{4}} x\cdot 2\cos 2x dx$, $D(X)=\int\limits_0^{\frac{\pi}{4}} x^2 2\cos 2x dx$;

4) $M(X)=\int\limits_0^{\frac{\pi}{4}} 2\cos 2x dx$, $D(X)=\int\limits_0^{\frac{\pi}{4}} x^2 2\cos 2x dx-\left(M(x)\right)^2$.

Тест 3.12. Математическое ожидание непрерывной случайной величины, заданной функцией распределения:

$$F(x)=\begin{cases}0, \textit{при}\, x\le 0;\\ \sin 2x;\textit{при}\, 0<x\le\frac{\pi}{4};\\ 1;\textit{при}\, x>\frac{\pi}{4}.\end{cases}$$

равно:

1) $M(X)=2\int\limits_0^{\frac{\pi}{4}} x\cos 2x dx$;

2) $M(X)=\int\limits_0^{\frac{\pi}{4}} \sin 2x dx$;

3) $M(X)=\int\limits_0^{\frac{\pi}{4}} x\sin 2x dx$;

4) $M(X)=\int\limits_0^{\frac{\pi}{4}} \cos 2x dx$;

5) $M(X)=2\int\limits_0^{\frac{\pi}{4}} \cos 2x dx$.

Тест 3.13. Дисперсия непрерывной случайной величины, заданной функцией распределения:

$$F(x)=\begin{cases}0,\, при\, x\le 0;\\ \sin 2x;\, при\, 0<x\le\dfrac{\pi}{4};\\ 1;\, при\, x>\dfrac{\pi}{4}.\end{cases}$$

равна:

1) $Д(X)=2\int\limits_0^{\frac{\pi}{4}} x\cos 2x dx$;

2) $Д(X)=\int\limits_0^{\frac{\pi}{4}} x\sin 2x dx-\left(M\left(x\right)\right)^2$;

3) $Д(X)=\int\limits_0^{\frac{\pi}{4}} x\cdot 2\cos 2x dx-\left(M\left(x\right)\right)^2$;

4) $Д(X)=\int\limits_0^{\frac{\pi}{4}} x\cos 2x dx-\left(M\left(x\right)\right)^2$;

5) $Д(X)=\int\limits_0^{\frac{\pi}{4}} 2x\cos 2x dx-\left(M\left(x\right)\right)$.

Тест 3.14. Оценку среднего значения случайной величины дает:
1) математическое ожидание;
2) дисперсия;
3) функция распределения;
4) плотность распределения вероятностей (дифференциальная функция распределения).

Тест 3.15. Степень разброса значений случайной величины вокруг ее математического ожидания определяет:
1) математическое ожидание;
2) дисперсия;
3) функция распределения;
4) плотность распределения вероятностей (дифференциальная функция распределения).

Вопросы для самоконтроля

1. Понятия случайной величины, дискретной случайной величины, непрерывной случайной величины.
2. Понятие закона распределения. Законы распределения дискретной случайной величины
3. Числовые характеристики дискретной случайной величины.
4. Понятие непрерывной случайной величины.
5. Понятие плотности распределения (дифференциальной функции) непрерывной случайной величины.
6. Числовые характеристики непрерывной случайной величины.
7. Вероятность попадания непрерывной случайной величины X через $f(x); F(x)$.

Ответы на тестовые задания

Номер теста	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9
Правильный ответ	1	2	4	5	5	4	2	5	4

Номер теста	3.10	3.11	3.12	3.13	3.14	3.15
Правильный ответ	2	1	1	3	1	2

4. НЕКОТОРЫЕ ЗАКОНЫ РАСПРЕДЕЛЕНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН

4.1. Биномиальный закон распределения

Случайная величина X, которая может принять возможное значение $X = k$ ($k = 0; 1; \ldots; n$) с вероятностью, определяемой по формуле Бернулли

$$P(X = k) = C_n^k p^k q^{n-k} \;,$$

называется *распределенной по биномиальному закону.*

Постоянные n и p ($q = 1 - p$) называются параметрами биномиального распределения.

Теорема. Числовые характеристики случайной величины, распределенной по биномиальному закону, вычисляются по формулам:

$$M(X) = np\,;\; D(X) = npq\,.$$

Таким образом, в примере 3.9 математическое ожидание и дисперсию можно было вычислять следующим образом:

$$M(X) = 4\cdot\frac{1}{2} = 2\,;\quad D(X) = 4\cdot\frac{1}{2}\cdot\frac{1}{2} = 1,$$

здесь $n = 4$; $p = \dfrac{1}{2}$; $q = \dfrac{1}{2}$.

Тест 4.1. Монету бросают 4 раза. Случайная величина X – число выпадений герба. Математическое ожидание случайной величины X равно:

1) 2;
2) 0;
3) 1;
4) $\dfrac{1}{4}$;
5) 4.

Тест 4.2. Монету бросают 4 раза. Случайная величина X – число выпадений герба. Дисперсия случайной величины X равна:

1) 2;
2) 0;
3) 1;
4) $\dfrac{1}{4}$;
5) 4.

Тест 4.3. Случайная величина X называется распределенной по биномиальному закону, если она принимает возможные значения с вероятностью, определяемой по формуле:

1) $P(X = m) = C_n^m p^m q^{n-m}$;
2) $P(X = m) = \dfrac{\lambda^m e^{-\lambda}}{m!}$;
3) $f(x) = \begin{cases} 0, \; где\, x \le 0; \\ ke^{-kx}, \; где\, x > 0, k > 0 \end{cases}$;
4) $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$;
5) $f(x) = \begin{cases} C, \quad где\, a \le x \le b; \\ 0, \quad где\, x > b,\, x < a \end{cases}$.

Тест 4.4. Математическое ожидание, дисперсия непрерывной случайной величины X, биномиально распределенной случайной величины равны:

1) n; np^2;

2) np, npq;

3) $\frac{p}{n}$; p;

4) $\frac{n}{p}$; $p(1-p)$;

5) p, p^2.

4.2. Закон Пуассона

Случайная величина X, которая может принять возможное значение $X = k$ ($k = 0;1;\ \ldots$) с вероятностью, определяемой по формуле Пуассона

$$P(X = k) = \frac{\lambda^k \cdot e^{-\lambda}}{k!},$$

называется *распределенной по закону Пуассона*.

Постоянная $\lambda = np$ называется параметром.

Теорема. Математическое ожидание и дисперсия случайной величины X, распределенной по закону Пуассона равны параметру λ.

Тест 4.5. Математическое ожидание случайной величины X, распределенной по закону Пуассона:

$$\mathrm{P}_n(m) = \frac{\lambda^m \cdot e^{-\lambda}}{m!},$$

равно:
1) m;
2) λ;
3) λ^2;
4) λ^m;
5) $e^{-\lambda}$.

Тест 4.6. Дисперсия случайной величины X, распределенной по закону Пуассона:

$$\mathrm{P}_n(m) = \frac{\lambda^m \cdot e^{-\lambda}}{m!},$$

равна:
1) m;
2) λ;
3) λ^2;
4) λ^m;
5) $e^{-\lambda}$.

Тест 4.7. Случайная величина X называется распределенной по закону Пуассона, если она принимает возможные значения с вероятностью, определяемой по формуле:

1) $P(X = m) = C_n^m p^m q^{n-m}$;

2) $P(X = m) = \frac{\lambda^m e^{-\lambda}}{m!}$;

3) $f(x) = \begin{cases} 0, & где\, x \le 0; \\ ke^{-kx}, & где\, x > 0,\, k > 0 \end{cases}$;

4) $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$;

5) $f(x) = \begin{cases} C, & где\, a \le x \le b; \\ 0, & где\, x > b,\, x < a \end{cases}$.

4.3. Равномерное распределение

Непрерывная случайная величина называется *равномерно распределенной на интервале* $(a;b)$, если плотность ее распределения имеет следующий вид:

$$f(x)=\begin{cases}0;\ x\notin(a;b),\\ \dfrac{1}{b-a};\ x\in(a;b).\end{cases}$$

Функция распределения равномерно распределенной случайной величины имеет следующий вид:

$$F(x)=\begin{cases}0,\ x\le a,\\ \dfrac{x-a}{b-a},\ a<x\le b,\\ 1,\ x>b.\end{cases}$$

Теорема. Для равномерно распределенной случайной величины математическое ожидание вычисляется по формуле $M(X)=\dfrac{a+b}{2}$, дисперсия вычисляется по формуле $D(X)=\dfrac{(b-a)^2}{12}$, среднее квадратическое отклонение вычисляется по формуле $\sigma(X)=\dfrac{b-a}{2\sqrt{3}}$.

Пример 4.1. Найти математическое ожидание, дисперсию, среднее квадратическое отклонение случайной величины *X*, распределенной равномерно в интервале $(2;8)$.

Решение

1. Математическое ожидание находим по формуле $M(X)=\dfrac{a+b}{2}$. По условию $a=2$; $b=8$. Следовательно, имеем:

$$M(X)=\frac{2+8}{2}=5.$$

2. Дисперсию находим по формуле $D(X)=\dfrac{(b-a)^2}{12}$.

Имеем: $D(X)=\dfrac{(8-2)^2}{12}=\dfrac{36}{12}=3$.

3. Находим среднее квадратическое отклонение:

$$\sigma(X)=\frac{8-2}{2\sqrt{3}}=\frac{6}{2\sqrt{3}}=\frac{3}{\sqrt{3}}=\sqrt{3}.$$

Примечание. Математическое ожидание, дисперсию, среднее квадратическое отклонение можно находить также по определению.

Тест 4.8. Среднее квадратическое отклонение случайной величины *X* распределенной равномерно на интервале (2;8), равно:

1) $\sqrt{3}$;
2) 0;
3) 1;
4) $\dfrac{1}{4}$;
5) 4.

Тест 4.9. Дисперсия случайной величины *X*, распределенной равномерно на интервале (2;8), равна:

1) 3;
2) 0;
3) 1;
4) $\dfrac{1}{4}$;
5) 5.

Тест 4.10. Математическое ожидание случайной величины X, распределенной равномерно на интервале (2;8), равно:

1) 2;
2) 0;
3) 1;
4) $\sqrt{3}$;
5) 5.

Тест 4.11. Случайная величина X называется равномерно распределенной на интервале $(a;b)$, если ее плотность распределения имеет вид:

1) $P(X=m)=C_n^m p^m q^{n-m}$;
2) $P(X=m)=\dfrac{\lambda^m e^{-\lambda}}{m!}$;
3) $f(x)=\begin{cases}0, & где\, x\le 0;\\ ke^{-kx}, & где\, x>0,\, k>0\end{cases}$;
4) $f(x)=\dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$;
5) $f(x)=\begin{cases}C, & где\, a\le x\le b;\\ 0, & где\, x>b,\, x<a\end{cases}$.

Тест 4.12. Если случайная величина подчинена закону равномерного распределения на интервале $(0;4)$, ее плотность распределения равна:

1) $f(x)=\begin{cases}1, & где\, 0\le x\le 4;\\ 0, & где\, x>4,\, x<0\end{cases}$;
2) $f(x)=\begin{cases}4, & где\, 0\le x\le 4;\\ 0, & где\, x>4,\, x<0\end{cases}$;
3) $f(x)=\begin{cases}0, & где\, 0\le x\le 4;\\ \dfrac{1}{4}, & где\, x>4,\, x<0\end{cases}$;
4) $f(x)=\begin{cases}\dfrac{1}{4}, & где\, 0\le x\le 4;\\ 0, & где\, x>4,\, x<0\end{cases}$.

Тест 4.13. Закон равномерного распределения задан дифференциальной функцией $f(x)=\dfrac{1}{b-a}$ в интервале $(a;b)$ и $f(x)=0$ вне этого интервала. Интегральная функция $F(X)$ на интервале $(a;b)$ будет равна:

1) $F(X)=0$;
2) $F(X)=\dfrac{x-a}{b-a}$;
3) $F(X)=\dfrac{x-b}{b-a}$;
4) $F(X)=1$;
5) $F(X)=\dfrac{b-a}{2}$.

Тест 4.14. Если, случайная величина подчинена закону равномерного распределения на интервале $(0,4)$, ее математическое ожидание равно:

1) $\dfrac{0+4}{2}$;
2) $\dfrac{0-4}{2}$;
3) $\dfrac{4}{0-4}$;
4) $\dfrac{1}{4-0}$.

4.4. Показательное распределение

Если дифференциальная функция (плотность) распределения вероятностей случайной величины X выражается функцией

$$f(x) = \begin{cases} 0, при\, x \le 0, \\ k \cdot e^{-kx}, при\, x > 0, \end{cases} \qquad (4.1)$$

где $k > 0$ – параметр, то говорят, что случайная величина X имеет *показательное распределение.*

Функция распределения такой случайной величины имеет вид

$$F(X) = \begin{cases} 0, при\, x \le 0, \\ 1 - e^{-kx}, при\, x > 0. \end{cases} \qquad (4.2)$$

Теорема. Числовые характеристики случайной величины, распределенной по показательному закону, определяются по формулам:

$$M(X) = \frac{1}{k};\; D(X) = \frac{1}{k^2};\; \sigma(X) = \frac{1}{k}.$$

Пример 4.2. Написать дифференциальную и интегральную функции показательно распределенной случайной величины X, если параметр $k = 6$.

Решение

Подставив k в соотношения (4.1) и (4.2), получим

$$f(x) = \begin{cases} 0, при\, x \le 0, \\ 6e^{-6x}, при\, x > 0; \end{cases}$$

$$F(x) = \begin{cases} 0, при\, x \le 0, \\ 1 - e^{-6x}, при\, x > 0. \end{cases}$$

Пример 4.3. Непрерывная случайная величина X распределена по показательному закону, заданному интегральной функцией

$$F(x) = \begin{cases} 0, при\, x \le 0, \\ 1 - e^{-0,6x}, при\, x > 0. \end{cases}$$

Найти вероятность того, что в результате испытания случайная величина попадет в интервал (2; 5).

Решение

Воспользуемся формулами вероятности попадания в интервал (a; b) случайной величины X:

1. Через $F(x)$:

$$P(2 < X < 5) = F(5) - F(2) = (1 - e^{-0,6\cdot 5}) - (1 - e^{-0,6\cdot 2}) = e^{-1,2} - e^{-3}.$$

2. Через $f(x)$

Найдем $f(x)$:

$$f(x) = F'(x) = \begin{cases} 0, при\, x \le 0, \\ 0,6e^{-0,6x}, при\, x > 0. \end{cases}$$

$$\text{Тогда } P(2 < X < 5) = \int_2^5 0,6 \cdot e^{-0,6x} dx = 0,6 \int_2^5 e^{-0,6x} dx = 0,6\left(-\frac{1}{0,6} e^{-0,6}\right) \Big|_2^5 =$$
$$= e^{-1,2} - e^{-3}$$

Ответ: $e^{-1,2} - e^{-3}$.

Пример 4.4. Найти математическое ожидание, дисперсию и сред-нее квадратическое отклонение показательного распределения, заданного дифференциальной функцией

$$f(x) = \begin{cases} 0, при\, x \le 0, \\ 5e^{-5x}, при\, x > 0. \end{cases}$$

Решение

Подставив $k = 5$ в формулы для вычисления математического ожидания, дисперсии и среднего квадратического отклонения, получим:

$$M(X)=\frac{1}{5},\ D(X)=\frac{1}{5^2},\ \sigma(X)=\frac{1}{5}.$$

Тест 4.15. Параметр *k* показательного распределения, заданного дифференциальной функцией:

$$f(x)=\begin{cases}0, при\, x\le 0,\\ 4\cdot e^{-4x}, при\, x>0,\end{cases}$$

равен:
1) 2;
2) 0;
3) 1;
4) $\frac{1}{4}$;
5) 4.

Тест 4.16. Дисперсия показательного распределения, заданного дифференциальной функцией:

$$f(x)=\begin{cases}0, при\, x\le 0,\\ 5\cdot e^{-5x}, при\, x>0,\end{cases}$$

равна:
1) $\frac{1}{25}$;
2) 5;
3) 1;
4) $\frac{1}{5}$;
5) –5.

Тест 4.17. Математическое ожидание, дисперсия непрерывной случайной величины *X*, распределенной по показательному закону

$$f(x)=\begin{cases}0,\ где\, x\le 0;\\ \lambda e^{-\lambda x},\ где\, x>0,\ \lambda>0,\end{cases}$$

равны:
1) λ, λ;
2) $\lambda, \frac{1}{\lambda}$;
3) $\frac{1}{\lambda}, \frac{1}{\lambda^2}$;
4) 1,0;
5) $\frac{1}{\lambda}, \lambda$.

Тест 4.18. Случайная величина *X* имеет показательное распределение, если ее дифференциальная функция (плотность) распределения равна:

1) $P(X=m)=C_n^m p^m q^{n-m}$;
2) $P(X=m)=\frac{\lambda^m e^{-\lambda}}{m!}$;
3) $f(x)=\begin{cases}0, & где\, x\le 0;\\ ke^{-kx}, & где\, x>0,\ k>0\end{cases}$;
4) $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$;
5) $f(x)=\begin{cases}C, & где\, a\le x\le b;\\ 0, & где\, x>b,\ x<a\end{cases}$.

4.5. Нормальное распределение

Нормальным называют распределение вероятностей непрерывной случайной величины X, если дифференциальная функция имеет вид

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-a)^2}{2\sigma^2}},$$

где a и σ – параметры,

a – математическое ожидание;

σ – среднее квадратическое отклонение.

Функция распределения $F(x)$ случайной величины X, распределенной по нормальному закону, имеет вид

$$F(x)=\int_{-\infty}^{x} f(t)dt=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{\frac{-(t-a)^2}{2\sigma^2}}dt.$$

Вероятность того, что X примет значение, принадлежащее интервалу $(\alpha;\ \beta)$, вычисляется по формуле

$$P(\alpha < X < \beta)=\Phi\left(\frac{\beta-a}{\sigma}\right)-\Phi\left(\frac{\alpha-a}{\sigma}\right),$$

где $\Phi(X)=\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dx$ – функция Лапласса.

Вероятность того, что абсолютная величина отклонения меньше положительного числа δ :

$$P(|X-a|<\delta)=2\Phi\left(\frac{\delta}{\sigma}\right).$$

Тест 4.19. Случайная величина X распределена по нормальному закону с параметрами $a=2$, $\sigma=1$. Тогда $P(|X-10|<3)$ равна:

1) $2\Phi(\frac{1}{10})$;

2) $2\Phi(\frac{3}{1})$;

3) $\Phi(\frac{2}{10})$;

4) $\Phi(\frac{1}{10})$;

5) $\Phi(1)$.

В частности, при $a=0$ справедливо равенство

$$P(|X|<\delta)=2\Phi\left(\frac{\delta}{\sigma}\right).$$

«Правило трех сигм». Вероятность того, что абсолютная величина отклонения нормально распределенной случайной величины будет меньше утроенного среднего квадратического отклонения, равна 0,9973:

$$P(|X-a|<3\sigma)=2\Phi(3)=0{,}9973.$$

Свойства функции $f(x)=\dfrac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-a)^2}{2\sigma^2}}$:

1. Область определения – вся числовая ось.
2. $f(x)>0$.
3. $\lim\limits_{x\to\infty} f(x)=0$, следовательно, ось OX является горизонтальной асимптотой графика функции.
4. Функция $f(x)$ имеет в точке $x=a$ максимум, равный $\dfrac{1}{\sigma\sqrt{2\pi}}$.
5. График функции $f(x)$ симметричен относительно прямой $x=a$.
6. График функции $f(x)$ в точках $x=a\pm\sigma$ имеет перегиб.

На основании перечисленных свойств график функции (нормальная кривая) имеет вид, представленный на рис. 9.

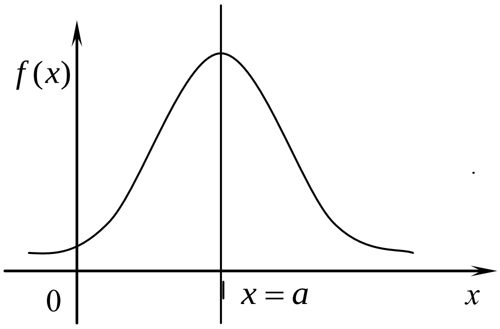

Рис. 9

Тест 4.20. На графике изображена кривая нормального распределения вероятностей (рис. 10).

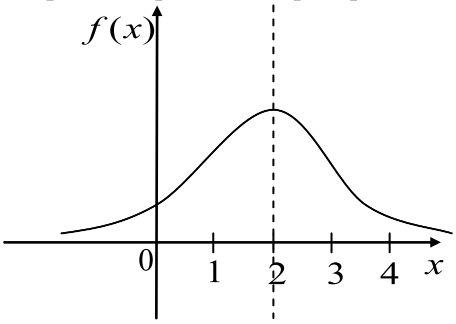

Рис. 10

Математическое ожидание равно:
1) $a = 1$;
2) $a = 2$;
3) a = 3;
4) $a = 4$;
5) $a = 0$.

Параметр a характеризует положение нормальной кривой, а параметр σ – форму.
На рис. 11 приведено положение нормальной кривой в зависимости от параметра a (если $a_1 < a_2$).

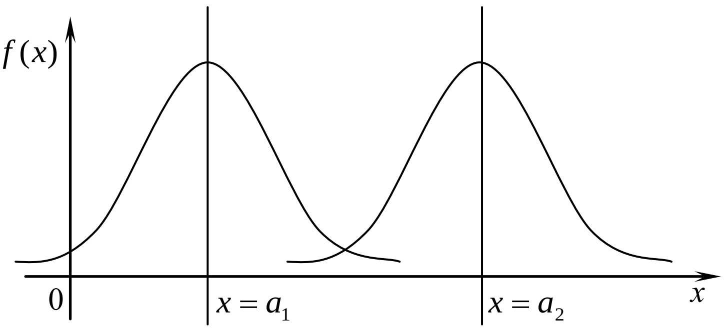

Рис. 11

На рис. 12 приведена форма нормальной кривой в зависимости от параметра σ (если $\sigma_1 < \sigma_2$).

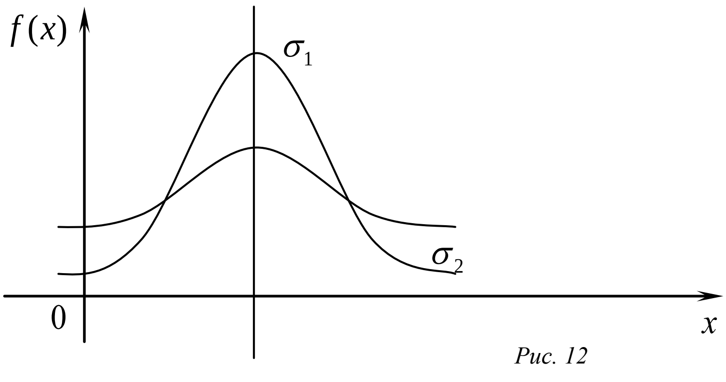

Рис. 12

Тест 4.21. На рисунке рис. 13 изображены три нормальные кривые.

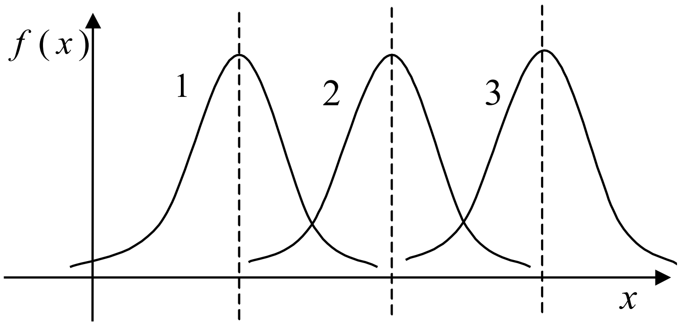

Рис. 13

Меньшему значению *a* соответствует кривая:

1) 1;
2) 2;
3) 3;
4) положение нормальной кривой не зависит от параметра *a*;
5) другой ответ.

Тест 4.22. На рис. 14 изображены три нормальные кривые.

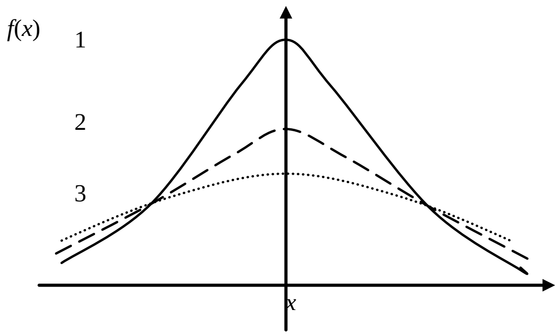

Рис. 14

Меньшему значению параметра σ соответствует нормальная кривая:

1) 1;
2) 2;
3) 3;
4) вид нормальной кривой не зависит от σ;
5) другой ответ.

Пример 4.5. Математическое ожидание и среднее квадратическое отклонение нормально распределенной случайной величины *X* соответственно равны 20 и 5. Какова вероятность того, что в результате испытания случайная величина *X* примет значение, заключенное в интервале (15;25).

Решение

Воспользуемся формулой

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right).$$

Подставив $\alpha = 15; \beta = 25; \sigma = 5$; a = 20, получим:

$$P(15 < X < 25) = \Phi\left(\frac{25-20}{5}\right) - \Phi\left(\frac{15-20}{5}\right) = \Phi(1) - \Phi(-1) =$$

$$(\textit{функция}\, \Phi(x) - \textit{нечетная}) = 2\Phi(1)$$

По таблице находим: $\Phi(x) = 0{,}3413$.

Таким образом, $P(15 < X < 25) = 2 \cdot 0{,}3413 = 0{,}6826$.

Ответ: 0,6826.

Пример 4.6. Случайная величина X распределена нормально с математическим ожиданием a = 15. Вероятность попадания X в интервал (15;20) равна 0,2. Чему равна вероятность попадания X в интервал (10;15)?

Решение

Так как график нормальной кривой симметричен относительно прямой $x = a = 15$,то площади, ограниченные сверху нормальной кривой и снизу интервалами (10;15) и (15;20) равны между собой (рис. 15).

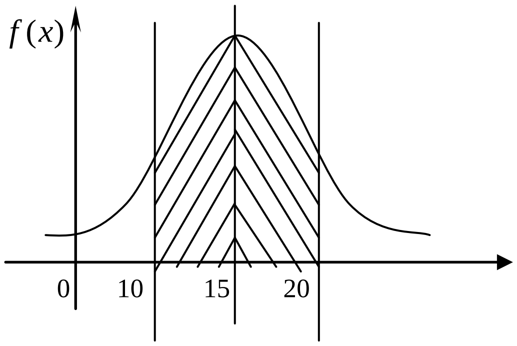

Рис. 15

Поскольку эти площади численно равны вероятностям попадания X в соответствующий интервал, то $P(10 < X < 15) = P(15 < X < 20) = 0{,}2$.

Ответ: 0,2.

Тест 4.23. Нормально распределенная случайная величина X задана дифференциальной функцией

$$f(x) = \frac{1}{5\sqrt{2\pi}} e^{\frac{-(x-1)^2}{50}},$$

Математическое ожидание равно:

1) –1;
2) 0;
3) 1;
4) 5;
5) $\frac{1}{50}$.

Тест 4.24. Нормально распределенная случайная величина X задана дифференциальной функцией

$$f(x) = \frac{1}{5\sqrt{2\pi}} e^{\frac{-(x-1)^2}{50}},$$

Среднее квадратическое отклонение равно:
1) –1;
2) 0;
3) 1;
4) 5;
5) $\frac{1}{50}$.

Тест 4.25. Случайная величина X распределена нормально с математическим ожиданием a = 8. Вероятность попадания X в интервал (0;4) равна 0,2. Вероятность попадания X в интервал (12;16) равна:
1) 0,1;
2) 0;
3)1;
4) 0,4;
5) 0,2.

Тест 4.26. Вероятность попадания в интервал (15;25) нормально распределенной случайной величины X с математическим ожиданием a = 20 и средним квадратическим отклонением $\sigma = 5$ равна:

1) $P(5 < X < 10) = \Phi\left(\frac{20-10}{5}\right) - \Phi\left(\frac{20-10}{5}\right)$;

2) $P(5 < X < 10) = \Phi\left(\frac{25-20}{5}\right) - \Phi\left(\frac{15-20}{5}\right)$;

3) $P(5 < X < 10) = \Phi(10-20) - \Phi(5-20)$.

Вопросы для самоконтроля

1. Биномиальный закон распределения случайной величин.
2. Закон распределения Пуассона.
3. Равномерный закон распределения.
4. Показательный закон распределения.
5. Нормальный закон распределения.

Ответы на тестовые задания

Номер теста	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	4.10
Правильный ответ	1	3	1	2	2	2	2	1	1	5

Номер теста	4.11	4.12	4.13	4.14	4.15	4.16	4.17	4.18	4.19	4.20
Правильный ответ	5	4	2	1	5	1	3	3	2	2

Номер теста	4.21	4.22	4.23	4.24	4.25	4.26
Правильный ответ	1	1	3	4	5	2

5. ДВУМЕРНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Двумерной случайной величиной называется случайная величина $(X;Y)$, задаваемая рядом распределения с двумя входами

y / x	y_1	…	y_m
x_1	p_{11}	…	p_{1m}
…	…	…	…
x_n	p_{n1}	…	p_{nm}

Так как события $\{x = x_i;\ y = y_j\}$ образуют полную группу, то $\sum_{i=1}^{n} p_{ij} = 1\ (j = 1, 2, ..., m)$.

Пример 5.1. Двумерная случайная величина задана рядом распределения.

y / x	$y_1 = 1$	$y_2 = 2$
$x_1 = 0$	0,1	0,4
$x_2 = 3$	0,2	$p(x_2;\ y_2)$

Какова вероятность $p(x_2;\ y_2)$?

Решение

$p(x_2;\ y_2) = 1 - (0,1 + 0,4 + 0,2) = 0,3$.

Ответ: 0,3.

Тест 5.1. Двумерная случайная величина задана рядом распределения

x / y	$y_1 = 1$	$y_2 = 2$
$x_1 = 0$	0,1	0,4
$x_2 = 3$	0,2	$p(x_2;\ y_2)$

Вероятность $p(x_2;\ y_2)$ равна:

1) 0;
2) 0,3;
3) 0,2;
4) 0,1;
5) 0,4.

Пример 5.2. Двумерная случайная величина задана рядом распределения

x \ y	$y_1 = 1$	$y_2 = 2$
$x_1 = 0$	0,1	0,4
$x_2 = 3$	0,2	0,3

Записать ряд распределения случайной величины X.

Решение

X	0	3
p_i	0,1 + 0,4	0,2 + 0,3

Тест 5.2. Двумерная дискретная величина (X, Y) задана законом распределения

x \ y	1	2
0	0,1	0,3
1	0,4	$p(x_2; y_2)$

Вероятность $p(x_2; y_2)$ равна:
1) 1;
2) 0,7;
3) 0,6;
4) 0,2;
5) 0.

Тест 5.3. Двумерная случайная величина задана рядом распределения

x \ y	$y_1 = 1$	$y_2 = 2$
$x_1 = 0$	0,1	0,4
$x_2 = 3$	0,2	0,3

Вероятность появления $x_2 = 3$ равна:
1) 0;
2) 0,1;
3) 0,2;
4) 0,3;
5) 0,5.

Тест 5.4. Двумерная случайная величина задана рядом распределения

x \ y	$y_1 = 1$	$y_2 = 2$
$x_1 = 0$	0,1	0,4
$x_2 = 3$	0,2	0,3

Вероятность появления $y_1 = 1$ равна:
1) 0;
2) 0,1;
3) 0,2;
4) 0,3;
5) 0,5.

Зависимость между случайными величинами x и y описывает *корреляционный момент:*

$$k_{xy} = M\big((X - M(X))(Y - M(Y))\big) = M(XY) - M(X)M(Y).$$

Коэффициентом корреляции r_{XY} случайных величин X и Y, между которыми предполагается линейная корреляционная связь, называется величина, определяемая по формуле

$$r_{XY} = \frac{M(XY) - M(X)\cdot M(Y)}{\sigma(X)\sigma(Y)}.$$

Корреляционной матрицей называется матрица вида

$$\begin{pmatrix} 1 & r_{XY} \\ r_{YX} & 1 \end{pmatrix}.$$

Корреляционная матрица также устанавливает взаимосвязь наборов выборочных данных по величине:

- при $0 < r_{XY} < 1$ большим значениям случайной величины X соответствуют большие значения случайной величины Y;
- при $-1 < r_{XY} < 0$ большим значениям случайной величины X соответствуют меньшие значения случайной величины Y (или наоборот);
- при $r_X = 0$ данные двух случайных величин некоррелированы;
- при $|r_{XY}| = 1$ существует линейная функциональная зависимость между случайными величинами X и Y.

Пример 5.3. Корреляционный момент $k_{xy} = 270$. Какова зависимость между X и Y?

Решение

Так как k_{xy} имеет размерность, то можно говорить лишь о прямой зависимости между X и Y.

Пример 5.4. Коэффициент корреляции $r_{XY} = -0{,}375$. Какова зависимость между X и Y?

Решение

По коэффициенту можно судить о виде зависимости и ее силе. Так как $r_{XY} = -0{,}375 < 0$, то зависимость обратная, так как $|r_{XY}| = 0{,}75$, то связь между X и Y высокая.

Тест 5.5. Известно, что $k_{xy} = 2{,}75$, $\sigma_X = 3{,}1$, $\sigma_Y = 2{,}5$. Коэффициент корреляции равен:

1) $r_{XY} = \dfrac{3{,}1\cdot 2{,}5}{2{,}75}$;
2) $r_{XY} = \dfrac{2{,}75}{3{,}1\cdot 2{,}5}$;
3) $r_{XY} = \dfrac{2{,}5}{3{,}1\cdot 2{,}75}$.

Тест 5.6. Коэффициент корреляции $r_{XY} = 0$. Тогда зависимость между X и Y:

1) прямая линейная;
2) обратная линейная;
3) данные двух случайных величин некоррелированы.

Тест 5.7. Коэффициент корреляции $r_{XY} = 1$. Тогда зависимость между X и Y:

1) прямая линейная;
2) обратная линейная;
3) данные двух случайных величин некоррелированы;
4) функциональная прямая линейная.

Тест 5.8. Коэффициент корреляции $r_{XY} = -1$. Тогда зависимость между X и Y:

1) прямая линейная;
2) обратная линейная;
3) данные двух случайных величин некоррелированы;
4) функциональная обратная линейная.

Вопросы для самоконтроля

1. Двумерные случайные величины.
2. Корреляционный момент.
3. Коэффициент корреляции.

Ответы на тестовые задания

Номер теста	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9
Правильный ответ	2	4	5	4	2	3	4	4	1

6. ЗАКОН БОЛЬШИХ ЧИСЕЛ

6.1. Неравенство Маркова

Теорема. Если случайная величина X может принимать только неотрицательные значения и у нее есть математическое ожидание, то какова бы ни была положительная величина ξ той же размерности, что и X, всегда выполняется неравенство

$$P(X < \xi) \ge 1 - \frac{M(X)}{\xi}. \qquad (6.1)$$

Известно, что сумма вероятностей двух противоположных событий равна 1, следовательно,

$$P(X < \xi) + P(X \ge \xi) = 1$$

или

$$P(X \ge \xi) = 1 - P(X < \xi).$$

Отсюда и из неравенства (6.1) следует, что

$$P(X \ge \xi) \le \frac{M(X)}{\xi}. \qquad (6.2)$$

Неравенства (6.1) и (6.2) называют *неравенствами Маркова.*

Тест 6.1. Чтобы имело место неравенство Маркова, случайная величина *X* должна:
1) принимать только неотрицательные значения;
2) иметь математическое ожидание;
3) принимать только положительные значения и иметь математи-ческое ожидание;
4) принимать только неотрицательные значения и иметь математическое ожидание;
5) принимать только отрицательные значения и иметь математи-ческое ожидание.

Пример 6.1. Средний срок службы мотора – 4 года. Используя неравенство Маркова, оценить вероятность того, что данный мотор прослужит менее 20 лет.

Решение

Пусть X – срок службы мотора. X – непрерывная случайная величина. Требуется оценить вероятность $P(X < 20)$. По условию $M(X) = 4$, $\xi = 20$. Воспользовавшись неравенством (6.1), получим $P(X < 20) \ge 1 - \frac{M(X)}{\xi} = 1 - \frac{4}{20} = 0{,}8$.

Ответ: $p \ge 0{,}8$.

Пример 6.2. Среднее число дождливых дней в году в данном пункте равно 90. Используя неравенство Маркова, оценить вероятность того, что в этом пункте будет не менее 150 дождливых дней в году.

Решение

Пусть *X* – число дождливых дней в году в данном пункте. *X* – дискретная случайная величина. По условию $M(X) = 90$, $\xi = 150$. Применяя неравенство (6.2), находим

$$P(X \ge 150) = \frac{M(X)}{\xi} = \frac{90}{150} = 0{,}6.$$

Ответ: $p \le 0{,}6$.

Тест 6.2. Средний вес клубня картофеля равен 120 г. Оценкой вероятности того, что наугад взятый клубень картофеля весит менее 360 г, по неравенству Маркова является выражение:

1) $1 - \frac{360}{120}$;

2) $1 - \frac{120}{360}$;

3) $\frac{120}{360}$;

4) $\frac{360}{120}$;

5) $1+\frac{120}{360}$.

Тест 6.3. Среднее число молодых специалистов, ежегодно направляемых в аспирантуру, составляет 200 человек. Оценкой вероятности того, что в данном году будет направлено в аспирантуру не менее 220 молодых специалистов, по неравенству Маркова является выражение:

1) $1+\frac{200}{220}$;

2) $\frac{220}{200}$;

3) $1-\frac{220}{200}$;

4) $1-\frac{200}{220}$;

5) $\frac{200}{220}$.

Тест 6.4. Среднесуточное потребление электроэнергии в населенном пункте равно 12000 кВт · ч. Оценкой вероятности того, что потребление электроэнергии в этом населенном пункте в течение данных суток не менее 50000 кВт·ч, по неравенству Маркова является выражение:

1) $\frac{50000}{12000}$;

2) $1-\frac{12000}{50000}$;

3) $\frac{12000}{50000}$;

4) $1-\frac{50000}{12000}$;

5) $\frac{12000}{50000}+\frac{50000}{12000}$.

6.2. Неравенство Чебышева

Теорема. Каково бы ни было $\varepsilon > 0$ для любой случайной величины X, дисперсия которой конечна, имеет место неравенство Чебышева

$$P(|X-M(X)|<\varepsilon)\geq 1-\frac{D(X)}{\varepsilon^2}\ . \qquad (6.3)$$

Известно, что сумма вероятностей двух противоположных событий равна единице, следовательно,

$$P(|X-M(X)|<\varepsilon)+P(|X-M(X)|\geq\varepsilon)=1$$

или

$$P(|X-M(X)|\geq\varepsilon)=1-P(|X-M(X)|<\varepsilon).$$

Отсюда и из неравенства (6.3) следует, что

$$P(|X-M(X)|\geq\varepsilon)\leq\frac{D(X)}{\varepsilon^2}\ . \qquad (6.4)$$

Неравенство (6.4) называют второй формой неравенства Чебышева.

Тест 6.5. Чтобы для случайной величины X имело место неравенство Чебышева, она должна обладать свойством:

1) математическое ожидание $M(X)$ конечно;

2) дисперсия $D(X)$ конечна;

3) $M(X)$ не существует;

4) $D(X) = \infty$;

5) $M(X) = \infty$.

Пример 6.3. Случайная величина X имеет дисперсию $D(X) = 1{,}66$. Используя неравенство Чебышева, оценить вероятность того, что случайная величина X отличается от своего математического ожидания по абсолютной величине менее чем на 2.

Решение

По условию $D(X)$ = 1,66, ε = 2. Применяя неравенство (6.3), находим $P(|X-M(X)|<2)\geq 1-\frac{1,66}{2^2}=0,585.$

Ответ: $p\geq 0,585.$

Пример 6.4. Случайная величина X имеет дисперсию $D(X)$ = 0,001.

Используя неравенство Чебышева, оценить вероятность того, что случайная величина X отличается от своего математического ожидания по абсолютной величине не менее чем на 0,1.

Решение

По условию $D(X)$ = 0,001, ε = 0,1.

Применяя неравенство (6.4), находим

$$P(|X-M(X)|\geq 0,1)\leq \frac{0,001}{(0,1)^2}=\frac{0,001}{0,01}=0,1.$$

Ответ: $p\leq 0,1$.

Тест 6.6. Случайная величина X имеет дисперсию $D(X)$ = 1620. Оценкой вероятности того, что случайная величина X отличается от своего математического ожидания по абсолютной величине менее чем на 200, по неравенству Чебышева является выражение:

1) $1+\frac{1620}{200^2}$;

2) $\frac{1620}{200^2}$;

3) $1-\frac{1620}{200^2}$;

4) $\frac{200^2}{1620}$;

5) $1-\frac{1620}{200}$.

Тест 6.7. Случайная величина X имеет дисперсию $D(X)$ = 0,004. Оценкой вероятности того, что случайная величина X отличается от своего математического ожидания по абсолютной величине не менее чем на 0,2, по неравенству Чебышева является выражение:

1) $\frac{0,004}{0,2}$;

2) $1-\frac{0,004}{0,2}$;

3) $1-\frac{0,004}{(0,2)^2}$;

4) $\frac{0,004}{(0,2)^2}$;

5) $\frac{(0,2)^2}{0,004}$.

6.3. Теорема Чебышева

Заранее нельзя уверенно предвидеть, какое из возможных значений примет случайная величина в итоге испытания; это зависит от многих случайных причин, учесть которые невозможно. Казалось бы, поскольку о каждой случайной величине мы располагаем в этом смысле весьма скромными сведениями, то вряд ли можно установить закономерности поведения и суммы достаточно большого числа случайных величин. На самом деле это не так. При некоторых сравнительно широких условиях суммарное поведение достаточно большого числа случайных величин почти утрачивает случайный характер и становится закономерным.

Для практики очень важно знание условий, при выполнении которых совокупное действие очень многих случайных причин приводит к результату, почти не зависящему от случая, так как позволяет предвидеть ход явлений. Эти условия и указываются в теоремах, носящих общее название закона больших чисел. К ним относятся теоремы Чебышева и Бернулли.

Теорема Чебышева. Если $X_1, X_2, \ldots, X_n, \ldots$ – последовательность независимых случайных величин, у каждой из которых есть математическое ожидание $M(X_i) = a_i$ и дисперсия $D(X_i) = \sigma_i^2$ $(i = 1, 2, \ldots, n, \ldots)$, причем дисперсии ограничены сверху одним и тем же числом C, то для любого положительного числа ε выполняется неравенство

$$P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \frac{a_1 + a_2 + \ldots + a_n}{n}\right| < \varepsilon\right) \ge 1 - \frac{C}{n\varepsilon^2}. \quad (6.5)$$

Данная теорема называется *законом больших чисел в форме Чебышева.*

Следствие 1. В неравенстве (6.5) перейдем к пределу при $n \to \infty$, получим

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \frac{a_1 + a_2 + \ldots + a_n}{n}\right| < \varepsilon\right) \ge 1.$$

Но так как вероятность больше единицы не бывает, то

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \frac{a_1 + a_2 + \ldots + a_n}{n}\right| < \varepsilon\right) = 1. \quad (6.6)$$

Следствие 2. Если $X_1, X_2, \ldots, X_n, \ldots$ – последовательность независимых случайных величин, математические ожидания каждой из которых равны a, а дисперсии – σ^2, то неравенство (6.5), формула (6.6) соответственно принимают вид

$$P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - a\right| < \varepsilon\right) \ge 1 - \frac{\sigma^2}{n\varepsilon^2},$$

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - a\right| < \varepsilon\right) = 1.$$

Данные результаты имеют большое практическое применение. Так, если надо измерить некоторую величину, истинное значение которой равно a, проводят n измерений этой величины. Результат каждого измерения – это случайная величина X_i ($i = 1,2,\ldots,n,\ldots$). Если при измерениях отсутствует систематическая погрешность, то $M(X_i) = a$. Можно считать, что X_i – независимая случайная величина, а ее дисперсия ограничена. Тогда справедливо следствие, и, значит, среднее арифметическое значение результатов измерений с ростом n приближается к истинному значению измеряемой величины a. Поэтому можно положить

$$\frac{X_1 + X_2 + \ldots + X_n}{n} \approx a.$$

Тест 6.8. Случайные величины $X_1, X_2, \ldots, X_n, \ldots$, для которых применим закон больших чисел в форме Чебышева должны иметь следующие числовые характеристики:

1) $M(X_i)$ и $D(X_i)$ ($i = 1,2,\ldots,n,\ldots$), причем математические ожидания ограничены сверху одним и тем же числом C;

2) $M(X_i)$ ($i = 1,2,\ldots,n,\ldots$), ограниченные сверху одним и тем же числом C;

3) $M(X_i)$ и $D(X_i)$ ($i = 1,2,\ldots,n,\ldots$);

4) $M(X_i)$ и $D(X_i)$ ($i = 1,2,\ldots,n,\ldots$), причем дисперсии ограничены сверху одним и тем же числом C;

5) $D(X_i)$ ($i = 1,2,\ldots,n,\ldots$).

Пример 6.5. Для определения средней продолжительности горения электроламп в партии из 200 одинаковых ящиков было взято на выборку по одной лампе из каждого ящика. Используя терему Чебышева, оценить вероятность того, что средняя продолжительность горения отобранных 200 электроламп отличается от средней продолжительности горения лампы во всей партии по абсолютной величине меньше, чем на 5 ч, если известно, что дисперсия продолжительности горения любой лампы в каждом ящике меньше 49.

Решение

Пусть X_i – продолжительность горения электролампы, взятой из i-го ящика. По условию $D(X_i) < 49$. Очевидно, что средняя продолжительность горения лампы в выборке $\dfrac{X_1 + X_2 + \ldots + X_{200}}{200}$, а средняя продолжительность горения ламп во всей партии $\dfrac{M(X_1) + M(X_2) + \ldots + M(X_{200})}{200}$.

Оценим сразу вероятность

$$P\left(\left|\frac{X_1+X_2+...+X_{200}}{200}-\frac{M(X_1)+M(X_2)+...M(X_{200})}{200}\right|<5\right).$$

Это есть левая часть неравенства (6.5). Так как $X_1, X_2, ..., X_{200}$ – независимые случайные величины, то эта вероятность оценивается правой частью неравенства (6.5), где следует положить C = 49, ε = 5, n = 200.

Итак, искомая вероятность

$$P \geq 1-\frac{49}{200\cdot 25}=1-\frac{49}{5000}=1-0{,}0098=0{,}9902.$$

Ответ: $p \geq 0{,}9902$.

Тест 6.9. За значение некоторой величины принимают среднее арифметическое достаточно большого числа ее измерений. Известно, что дисперсия возможных результатов каждого измерения не превосходит 25 мм. Оценкой вероятности того, что при 1000 измерениях неизвестной величины отклонение принятого значения от истинного по абсолютной величине не превзойдет 0,5 мм, по теореме Чебышева является выражение:

1) $1-\dfrac{25^2}{1000\cdot 0{,}5}$;

2) $1-\dfrac{0{,}5}{1000\cdot (0{,}5)^2}$;

3) $1-\dfrac{25}{1000\cdot (0{,}5)^2}$;

4) $1+\dfrac{25}{1000\cdot (0{,}5)^2}$;

5) $\dfrac{25}{1000\cdot (0{,}5)^2}$.

Тест 6.10. Для определения средней урожайности поля площадью 1800 га взяли на выборку 1 м^2 с каждого гектара. Известно, что по каждому гектару поля дисперсия не превышает 6. Оценкой вероятности того, что отклонение средней выборочной урожайности отличается от средней урожайности по всему полю по абсолютной величине меньше, чем на 0,25 ц, по теореме Чебышева является выражение:

1) $1-\dfrac{6}{1800\cdot (0{,}25)^2}$;

2) $1+\dfrac{6^2}{1800\cdot (0{,}25)^2}$;

3) $\dfrac{6^2}{1800\cdot (0{,}25)^2}$;

4) $1-\dfrac{6^2}{1800\cdot (0{,}25)^2}$;

5) $1-\dfrac{1800\cdot (0{,}25)^2}{6}$.

6.4. Теорема Бернулли

Теорема Бернулли. Если в каждом из n независимых испытаний вероятность p появления события A постоянна, то вероятность того, что отклонение частости от вероятности p меньше по абсолютной величине положительного числа ε, не меньше, чем разность $1-\dfrac{pg}{n\varepsilon^2}$ $(q=1-p)$, т. е.

$$P\left(\left|\frac{m}{n}-p\right|<\varepsilon\right)\geq 1-\frac{pq}{n\varepsilon^2}. \tag{6.7}$$

Данная теорема называется *законом больших чисел в форме Бернулли.*

Примечание. Эта теорема опубликована в 1713 г., она положила начало теории вероятностей как науке. Доказательство Я. Бернулли было очень сложным. Более простое доказательство предложил П. Л. Чебышев в 1846 г.

Тест 6.11. Вероятность появления события A в n независимых испытаниях при применении закона больших чисел в форме Бернулли должна быть:

1) $p > 0{,}5$;
2) произвольная;
3) $p < 0{,}5$;
4) постоянная;
5) $p \to 0$.

Пример 6.7. Вероятность положительного исхода отдельного испытания $p = 0{,}8$. Оценить вероятность того, что в 1000 независимых повторных испытаниях отклонение частости положительных исходов от вероятности при отдельном испытании по абсолютной величине будет меньше 0,05.

Решение

В данном случае $n = 1000$; $p = 0{,}8$; $\varepsilon = 0{,}05$. Применяя неравенство (6.7), имеем

$$P\left(\left|\frac{m}{n}-0{,}8\right|<0{,}05\right)\geq 1-\frac{0{,}8(1-0{,}8)}{1000\cdot(0{,}05)^2}=1-\frac{0{,}8\cdot 0{,}2}{2{,}5}=0{,}936.$$

Ответ: $p \geq 0{,}936$.

Пример 6.8. Оценить вероятность того, что частость появления грани с четным числом очков в 1000 независимых подбрасываниях игрального кубика отклоняется от вероятности появления грани с четным числом очков по абсолютной величине меньше чем на 0,1.

Решение

В данном случае $n = 1000$, $\varepsilon = 0{,}1$. Так как в отдельном подбрасывании игрального кубика вероятность выпадения грани с четным числом очков постоянна и равна $\frac{3}{6}=0{,}5$, то $p = 0{,}5$. Применяя неравенство (6.7), имеем

$$P\left(\left|\frac{m}{n}-0{,}5\right|<0{,}1\right)\geq 1-\frac{0{,}5\cdot(1-0{,}5)}{1000\cdot(0{,}1)^2}=1-\frac{0{,}5\cdot 0{,}5}{10}=0{,}975.$$

Ответ: $p \geq 0{,}975$.

Тест 6.12. Вероятность наступления события A в каждом испытании равна 0,4. Оценкой вероятности того, что в 20000 испытаний отклонение частости события A от его вероятности меньше по абсолютной величине 0,01, по теореме Бернулли является выражение:

1) $1-\frac{0{,}4\cdot 0{,}6}{20000\cdot 0{,}01}$;
2) $1-\frac{0{,}4}{20000\cdot(0{,}01)^2}$;
3) $1-\frac{0{,}4\cdot 0{,}5}{20000\cdot(0{,}01)^2}$;
4) $\frac{0{,}4\cdot 0{,}6}{20000\cdot(0{,}01)^2}$;
5) $1-\frac{0{,}4\cdot 0{,}6}{20000\cdot(0{,}01)^2}$.

Тест 6.13. Пусть вероятность того, что покупателю книжного магазина необходима книга *А*, равна 0,3. Оценкой вероятности того, что при 1000 побывавших в магазине покупателях отклонение частости покупки книги *А* от вероятности покупки этой книги меньше по абсолютной величине 0,145, по теореме Бернулли является выражение:

1) $1-\frac{0{,}3\cdot 0{,}7}{1000\cdot 0{,}145}$;

2) $1-\dfrac{0{,}3\cdot 0{,}7}{1000\cdot(0{,}145)^2}$;

3) $1+\dfrac{0{,}3\cdot 0{,}7}{1000\cdot(0{,}145)^2}$;

4) $\dfrac{0{,}3\cdot 0{,}7}{1000\cdot(0{,}145)^2}$;

5) $1-\dfrac{1000\cdot(0{,}145)^2}{0{,}3\cdot 0{,}7}$.

Тест 6.14. Оценкой вероятности того, что частность появления грани с шестью очками в 10000 независимых подбрасываниях игрального кубика отклоняется от вероятности появления грани с шестью очками по абсолютной величине меньше чем на 0,01, по теореме Бернулли является выражение:

1) $1-\dfrac{\frac{2}{3}\cdot\frac{1}{3}}{10000\cdot(0{,}01)^2}$;

2) $1-\dfrac{\frac{1}{6}}{10000\cdot(0{,}01)^2}$;

3) $1-\dfrac{\frac{1}{6}\cdot\frac{5}{6}}{10000\cdot(0{,}01)^2}$;

4) $1-\dfrac{\left(\frac{1}{6}\right)^2}{10000\cdot(0{,}01)^2}$;

5) $1-\dfrac{\left(\frac{1}{6}\cdot\frac{5}{6}\right)^2}{10000\cdot(0{,}01)^2}$.

Тест 6.15. Оценкой вероятности того, что частность появления герба в 100 независимых подбрасываниях монеты отклоняется от вероятности появления герба по абсолютной величине меньше чем на 0,01, по теореме Бернулли является выражение:

1) $1-\dfrac{\frac{1}{2}\cdot\frac{3}{2}}{100\cdot(0{,}01)^2}$;

2) $1-\dfrac{\frac{1}{2}}{100\cdot(0{,}01)^2}$;

3) $1-\dfrac{\frac{1}{3}\cdot\frac{2}{3}}{100\cdot 0{,}01}$;

4) $1-\dfrac{\frac{1}{2}\cdot\frac{1}{2}}{100\cdot(0{,}01)^2}$;

5) $\dfrac{\frac{1}{2}\cdot\frac{1}{2}}{100\cdot(0{,}01)^2}$.

Вопросы для самоконтроля

1. Запишите неравенство Маркова.
2. Запишите неравенство Чебышева.
3. Что называется законом больших чисел?

4. Сформулируйте теорему Чебышева.
5. Сформулируйте теорему Бернулли.
6. Закон больших чисел.

Ответы на тестовые задания

Номер теста	6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	6.10	6.11
Правильный ответ	4	2	5	3	2	3	4	4	1	3	4

Номер теста	6.12	6.13	6.14	6.15
Правильный ответ	5	2	3	4

7. ВЫБОРОЧНЫЙ МЕТОД

7. 1. Выборка

Пусть случайная величина X описывает количественный или качественный признак некоторого физического или экономического процесса. В реальных условиях обычно бывает трудно или экономически нецелесообразно, а иногда и невозможно, исследовать всю совокупность значений случайной величины (*генеральную совокупность*) исследуемого процесса. Поэтому на практике широко применяется выборочный метод, когда исследуется часть генеральной совокупности (*выборочная совокупность,* или *выборка*). При этом выборка должна быть *репрезентативной* (представительной), что, в силу закона больших чисел, достигается случайностью отбора. Различают пять основных типов выборок:

1. *Собственно-случайная:*
- *повторная* (элементы после выбора возвращаются обратно);
- *бесповторная* (выбранные элементы не возвращаются).

2. *Типическая* (генеральная совокупность предварительно разбивается на группы типических элементов, и выборка осуществляется из каждой):
- *равномерная* (при равенстве объемов исходных групп в генеральной совокупности выбирается одинаковое количество элементов из каждой);
- *пропорциональная* (численность выборок формируют пропорционально численностям или средним квадратическим отклонениям групп генеральной совокупности);
- *комбинированная* (численность выборок пропорциональна и средним квадратическим отклонениям, и численностям групп генеральной совокупности).

3. *Механическая* (отбор элементов проводится через определенный интервал).
4. *Серийная* (отбор проводится не по одному элементу, а сериями для проведения сплошного обследования).
5. *Комбинированная* – используются различные комбинации вышеуказанных методов, например, типическая выборка сочетается с механической и собственно случайной.

Тест 7.1. Вся совокупность объектов, характеризующая изучаемый признак, называется:
1) выборкой;
2) генеральной совокупностью;
3) величиной;
4) совокупностью.

Тест 7.2. Часть генеральной совокупности называется:
1) выборкой;
2) генеральной совокупностью;
3) величиной;
4) совокупностью.

Тест 7.3. Если элементы после выбора возвращаются обратно, то выборка:
1) повторная;
2) бесповторная;
3) типическая;
4) равномерная.

Тест 7.4. Если выбранные элементы не возвращаются, то выборка:
1) повторная;
2) бесповторная;
3) типическая;

4) равномерная.

7.2. Статистические ряды

Пусть из генеральной совокупности значений случайной величины X извлекается выборка $\{x_1; x_2; ...; x_n\}$ *объема n*. *Вариационным рядом* называется таблица, в *первой строке* которой указываются *варианты* $x_{(1)} = x_{min}$, $x_{(2)}$, …, $x_{(k)} = x_{max}$ наблюдаемого признака в порядке возрастания $x_{(1)} \le x_{(2)} \le ... \le x_{(k)}$, во второй строке – соответствующие им частоты n_i или относительные частоты $w_i\left(X = x_{(i)}\right) = \frac{n_i}{n}$ (табл. 1). Здесь n_i – количество наблюдений значения x_i в выборке, $\sum_{i=1}^{k} n_i = n$.

Таблица 1. **Вариационный ряд**

Варианты X	$x_{(1)}$	$x_{(2)}$	…	$x_{(k)}$
Частота	n_1	n_2	…	n_k
Относительные частоты w_i	$\frac{n_1}{n}$	$\frac{n_2}{n}$	…	$\frac{n_k}{n}$

Порядковый номер варианты в вариационном ряду называют ее *рангом*.

Тест 7.5. Даны значения признака X: 10, 5, 7, 4, 15. Ранг «10» равен?
1) 1;
2) 2;
3) 3;
4) 4;
5) 5.

Тест 7.6. Даны значения признака X: 13, 20, 15, 14, 21. Разность рангов «21» и «20» равна:
1) 1;
2) 2;
3) 3;
4) 4;
5) 5.

Тест 7.7. Даны значения признаков X и Y

X	2	13	20
Y	7	9	8

Произведение рангов $X = 20$ и $Y = 8$ равно:
1) 9;
2) 2;
3) 6;
4) 4;
5) 5.

Полигоном относительных частот называется ломаная линия, состоящая из отрезков, соединяющих точки $\left(x_{(i)}; \frac{n_i}{n}\right)$, $i = 1,2,...,k$.

Полигоном частот называется ломаная линия, состоящая из отрезков, соединяющих точки $\left(x_{(i)}; n_i\right)$, $i = 1,2,...,k$.

Тест 7.8. Ломаная, отрезки которой соединяют точки (x_1, n_1), (x_2, n_2), …, (x_k, n_k), где x_i – варианты выборки, n_i – соответствующие им частоты, называется:
1) полигоном относительных частот (частостей);
2) полигоном частот;
3) многоугольником распределения;
4) гистограммой.

При большом объеме выборки ее элементы могут быть сгруппированы в *интервальный вариационный ряд* (табл. 2).

Таблица 2. **Интервальный вариационный ряд**

Варианты X	$[x_{(1)}; x_{(1)}+h)$	$[x_{(1)}+h; x_{(1)}+2h)$	$\cdots$	$[x_{(1)}+(m-1)h; x_{(k)})$
Частота	n_1^*	n_2^*	$\cdots$	n_{m-1}^*
Относительные частоты w_i	$\frac{n_1^*}{n}$	$\frac{n_2^*}{n}$	$\cdots$	$\frac{n_{m-1}^*}{n}$

Для этого интервал изменения всех ее вариант разбивается на m непересекающихся полуинтервалов. Вычисления значительно упрощаются, если частичные интервалы имеют одинаковую длину $h=\frac{R}{m}$, где величина $R=x_{(k)}-x_{(1)}$ является *размахом* выборки. Затем подсчитывается число вариант выборки, попавших в каждый из интервалов, вычисляются относительные частоты числа вариант в интервале. Интервальный статистический ряд представляет собой таблицу, в *первой строке* которой указываются полуинтервалы $[x_{(1)}; x_{(1)}+h)$, $[x_{(1)}+h; x_{(1)}+2h)$, $\cdots$, $[x_{(1)}+(m-1)h; x_{(k)})$, *во второй строке* – соответствующие им частоты n_i^* или относительные частоты $w_i=\frac{n_i^*}{n}$, $i=1,2,...,m-1$, где n_i^* – количество наблюдений в соответствующем интервале, $\sum_{i=1}^{k} n_i^* = n$, $x_{(m)}=x_{(k)}$ – наибольшая варианта.

Если серединам $x_{(i)}^*$ каждого полуинтервала $[x_{(1)}+(i-1)h; x_{(1)}+ih)$ $(i=1,2,...,m-1)$ равной длины поставить в соответствие относительные частоты $w_i=\frac{n_i^*}{n}$, то получится *вариационный ряд с равноотстоящими вариантами* (табл. 3).

Таблица 3. **Вариационный ряд с равноотстоящими вариантами**

Варианты X	$x_{(1)}^*$	$x_{(2)}^*$	$\cdots$	$x_{(m-1)}^*$
Относительные частоты w_i	$\frac{n_1^*}{n}$	$\frac{n_2^*}{n}$	$\cdots$	$\frac{n_{m-1}^*}{n}$

Аналогично, для полуинтервалов разной длины получается вариационный ряд с *неравноотстоящими* вариантами.

Гистограммой относительных частот называется ступенчатая фигура, состоящая из прямоугольников, построенных на интервалах группировки $[x_{(i)}; x_{(i+1)})$, $i=1,2,...,m-1$, так, что площадь каждого прямоугольника равна относительной частоте. Основание каждого прямоугольника лежит на оси Ox и равно длине полуинтервала, высота равна величине $\frac{n_i^*}{n\cdot h}$. Сумма площадей всех прямоугольников равна единице.

По виду полигона или гистограммы обычно выдвигают предположение о виде закона распределения исследуемой случайной величины, что позволяет придать определенную направленность исследованиям.

7.3. Эмпирическая функция распределения и кумулятивная кривая

Статистической (эмпирической) функцией распределения выборки называется функция $F^*(x)$ действительного аргумента *x*, определяющая относительную частоту события {X < x}: $F^*(x)=\frac{n_x}{n}$, где n_x – число вариант, меньших *x*.

Основными свойствами статистической функции распределения выборки являются:

1) $0\le F^*(x)\le 1$;

2) функция $F^*(x)$ является неубывающей;

3) $F^*(x)=0$ при $x<x_1$ и $F^*(x)=1$ при $x>x_k$, где x_1 и x_k – соответственно наименьшее и наибольшее значения выборки;

4) $P(a<X<b)=F^*(b)-F^*(a)$.

Если известен вариационный ряд выборки, то статистическая функция распределения определяется как $F^*(x)=\frac{n_x}{n}=\sum\limits_{x_{(i)}<x}\frac{n_i}{n}$.

Для построения статистической функции распределения $F^*(x)$ в случае известного интервального ряда осуществляется переход к вариационному ряду с равноотстоящими или неравноотстоящими вариантами.

Пример 7.1. Приведена статистика по годовым темпам инфляции *X%* в некоторой стране за 10 лет: 1,8; 3,2; –0,6; 0,3; 1,8; –0,6; 2,8; 3,2; 5,1; 1,8.

Необходимо выполнить следующее:

1) построить вариационный ряд относительных частот;

2) построить полигон относительных частот;

3) построить эмпирическую функцию распределения $F^*(x)$;

4) оценить $P\,(1{,}7 < X < 2{,}9)$;

5) оценить $P(X > 3)$.

Решение

1. Сгруппируем данные и составим вариационный ряд:

x_i	-0,6	0,3	1,8	2,8	3,2	5,1
n_i	2	1	3	1	2	1
w_i	0,2	0,1	0,3	0,1	0,2	0,1

2. Полигон относительных частот представлен на рис. 16.

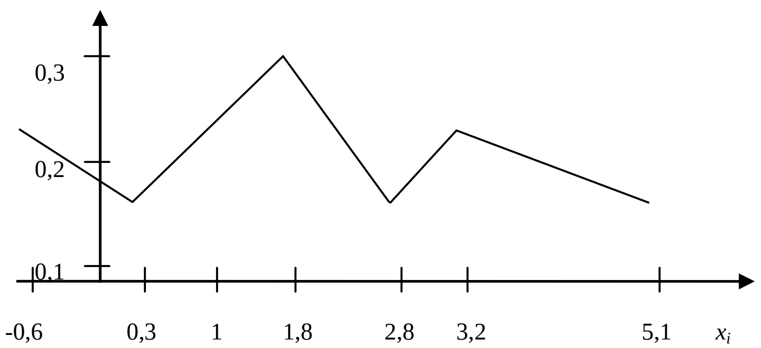

Рис. 16

3. Строим эмпирическую функцию распределения (по аналогии с функцией распределения дискретной величины):

$$F^*(x)=\begin{cases}0,\ \textit{если}\ x\le -0{,}6,\\ 0{,}2,\ \textit{если}\ -0{,}6<x\le 0{,}3,\\ 0{,}3,\ \textit{если}\ 0{,}3<x\le 1{,}8,\\ 0{,}6,\ \textit{если}\ 1{,}8<x\le 2{,}8,\\ 0{,}7,\ \textit{если}\ 2{,}8<x\le 3{,}2,\\ 0{,}9\ \textit{если}\ 3{,}2<x\le 5{,}1,\\ 1,\ \textit{если}\ x>5{,}1.\end{cases}$$

График этой функции изображена рис. 17.

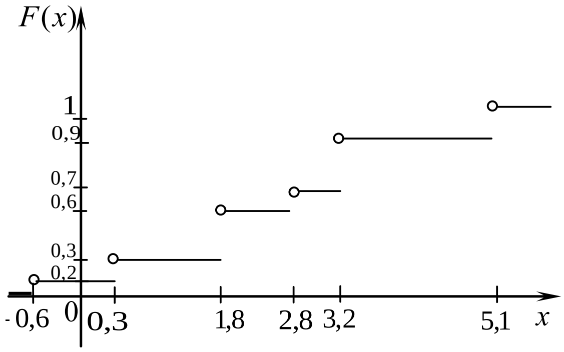

Рис. 17

4. P (1,7 < X< 2,9) = F(2,9) – F(1,7) = 0,8 – 0,6 = 0,2;
5. P(X > 3) = 1– P(X < 3) = 1– (0,2 + 0,1 + 0,3 + 0,1) = 1 – 0,7 = 0,3.

Тест 7.9. Задан вариационный ряд

X	1	3	4,5	6,2	7
$\frac{n_i}{n}$	0,2	0,15	0,4	?	0,05

Относительная частота варианты 6,2 равна:
1) 0,5;
2) 0,3;
3) 0,2;
4) 0,1.

Тест 7.10. Задан вариационный ряд

X	1	3	4,5	6,2	7
$\frac{n_i}{n}$	0,2	0,15	0,4	0,2	0,05

Ранг варианты 6,2 равен:
1) 5;
2) 4;
3) 2;
4) 1.

Тест 7. 11. Задан вариационный ряд

X	1	3	4,5	6,2	7
$\frac{n_i}{n}$	0,2	0,15	0,4	0,2	0,05

Вероятность $P(X > 5)$ равна:
1) 0,2;
2) 0,05;
3) 0,5;
4) 0,25.

Вопросы для самоконтроля

1. Что называется вариационным рядом?
2. Как строится интервальный ряд?
3. Какой ряд называется интервальным вариационным рядом с равноотстоящими вариантами?
4. Как строится гистограмма?
5. Что называется кумулятивной кривой?
6. Что называется статистической функцией распределения?
7. Какими свойствами обладает статистическая функция распределения?

7.4. Числовые характеристики выборки

Пусть случайный эксперимент описывается случайной величиной *X*, распределение которой зависит от одного или несколько *параметров*. К ним, в частности, относятся среднее, мода, медиана, среднее квадратичное отклонение, дисперсия, коэффициенты эксцесса и асимметрии, размах вариации, называемые *параметрами генеральной совокупности*. При исследовании случайной величины *X* из генеральной совокупности ее возможных значений извлекается выборка $X=\{x_1;x_2;...;x_n\}$ объема *n*. По данной выборке можно приближенно вычислить значения каждого из изучаемых параметров, которые в статистике называются *числовыми оценками параметров* или просто *оценками*.

Данные характеристики условно разбиваются на четыре группы:

- показатели положения вариант на числовой оси;
- показатели разброса вариант относительно своего центра, определяющие кучность данных около центра;
- показатели асимметрии распределения вариант около своего центра;
- показатели, описывающие закон распределения.

Пусть выборка задана вариационным рядом (табл. 4).

Таблица 4. **Вариационный ряд**

Варианты *X*	x_1	x_2	…	x_k
Относительные частоты w_i	$\frac{n_1}{n}$	$\frac{n_2}{n}$	…	$\frac{n_k}{n}$

Выборочным средним значением выборки называется число, определяемое по формуле $\bar{x}=\frac{1}{n}\sum_{i=1}^{k}x_i n_i$, где x_i – варианта с частотой n_i, *n* – число наблюдений, $\sum_{i=1}^{k}n_i=n$. Если частоты n_i равны единице, то $k = n$. При достаточно больших *n* используют формулу: $\bar{x}=\sum_{i=1}^{k}x_i w_i$, где *k* – число значений вариационного ряда, w_i – относительная частота варианты x_i.

Выборочной медианой Me называется значение признака, находящегося в середине вариационного ряда. Если число вариант нечетно, т. е. $n=2m+1$, то медианой является ($m+1$)-я варианта ($Me=x_{m+1}$); если же число вариант четно, то медиана равна среднему арифметическому двух значений в середине ряда: $Me=\frac{x_m+x_{m+1}}{2}$.

Выборочной модой (Mo) называется варианта выборки, имеющая наибольшую частоту. Если несколько соседних значений имеют наибольшую частоту, то модой является их среднее арифметическое: $Mo=\frac{x_i+x_{i+1}}{2}$, где для вариант x_i, x_{i+1} частоты $n_i=n_{i+1}=n_{\max}$. Если две или более несмежных вариант имеют разные наибольшие частоты, то ряд называется полимодальным. Если же все варианты встречаются одинаково часто, то ряд моды не имеет.

Мода и медиана используются в качестве характеристики среднего положения в случае, если границы ряда нечеткие или если ряд не симметричен.

Для описания рассеивания значений случайной величины относительно выборочного среднего используются выборочная дисперсия и выборочное среднее квадратическое отклонение.

Выборочной дисперсией значений выборки $\{x_1,x_2,\ldots,x_n\}$ называется число, определяемое по формуле

$$D_в=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2.$$

После преобразований получается следующая формула:

$$D_в=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\left(\bar{x}\right)^2.$$

Если задан вариационный ряд, то используется формула $\sigma_в^2=\sum_{i=1}^{k}(x_i-\bar{x})^2 w_i$, где *k* – число вариант, w_i – относительная частота варианты x_i.

Выборочным средним квадратическим отклонением называется число, которое находится по формуле $\sigma_в=\sqrt{D_в}$.

Пример 7.2. Выборка задана распределением частот

X	-1	0	1	2
n_i	1	3	2	4

Найти выборочную среднюю $\overline{X_в}$, выборочную дисперсию $D_в$, выборочное среднее квадратическое отклонение $\sigma_в$.

Решение

$\bar{x}_в = \frac{1}{10}\left(-1\cdot 1 + 0\cdot 3 + 1\cdot 2 + 2\cdot 4\right) = \frac{9}{10} = 0{,}9$;

$D_в = \frac{1}{10}\left((-1)^2\cdot 1 + 0^2\cdot 3 + 1^2\cdot 2 + 2^2\cdot 4\right) - (0{,}9)^2 = 1{,}09$;

$\sigma_в = \sqrt{1{,}09}$.

Тест 7. 12. Для вариационного ряда

X	-1	0	1
n_i	5	2	3

выборочное среднее $\overline{X_в}$, выборочная дисперсия $D_в$ равны:

1) $\overline{X_в} = 0$, $D_в = 1$;
2) $\overline{X_в} = 2$, $D_в = 10$;
3) $\overline{X_в} = -0{,}2$, $D_в = 0{,}76$;
4) $\overline{X_в} = 0{,}3$, $D_в = 0{,}8$.

Коэффициент вариации $\upsilon = \frac{\sigma_в}{\bar{x}}\cdot 100\%$ характеризует относительное значение среднего квадратического отклонения и служит для сравнения колеблемости несоизмеримых показателей.

Обобщающими характеристиками выборочных распределений являются моменты вариационного ряда.

Начальным выборочным моментом m-го порядка (m =0,1,2,…) называется величина $\nu_m^* = \frac{1}{n}\sum_{i=1}^{k} x_i^m n_i$, где x_i – наблюденное значение с частотой n_k, n – число наблюдений, $\sum_{i=1}^{k} n_i = n$. Очевидно, что $\nu_0^* = 1$, $\nu_1^* = \bar{x}$, $\nu_2^* = \overline{x^2}$ и т. д.

Центральным выборочным моментом m-го порядка (m =0,1,2,…) называется величина $\mu_m^* = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^m n_i$, где x_i – наблюденное значение с частотой n_i, n – число наблюдений, $\sum_{i=1}^{k} n_i = n$, $\bar{x}$ – выборочное среднее. Очевидно, что $\mu_0 = 1$, $\mu_1^* = 0$, $\mu_2^* = \sigma_в^2$,

$$\mu_3^* = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^3 n_i,\ \mu_4^* = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^4 n_i.$$

Важную роль при исследовании статистических совокупностей играют асимметрия и эксцесс распределения признака, которые вычисляются соответственно по формулам: $A_s = \frac{\mu_3^*}{\sigma_в^3}$, $E = \frac{\mu_4^*}{\sigma_в^4} - 3$.

Если кривая распределения симметрично относительно прямой $x = \bar{x}$, то распределение симметрично. Тогда $A_s = 0$ ($\mu_3^* = 0$). При асимметричном распределении вершина кривой сдвинута относительно ординаты выборочной средней. Если $A_s > 0$, то асимметрия правосторонняя или положительная (рис. 18а), если $A_s < 0$, то левосторонняя или отрицательная (рис. 18б).

а) б)

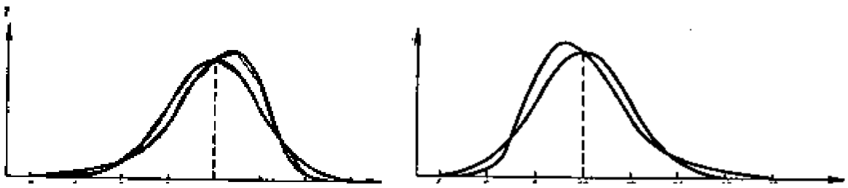

Рис. 18. **Правосторонняя (а) и левосторонняя (б) асимметричность распределения**

Эксцесс характеризует относительную остроконечность или сглаженность распределения по сравнению с нормальным распределением. Положительный эксцесс обозначает относительно остроконечное распределение (рис. 19а). Отрицательный эксцесс обозначает относительно сглаженное распределение (рис. 19б).

а) б)

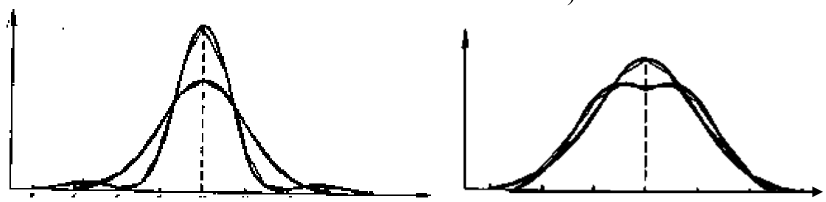

Рис. 19. **Положительный (а) и отрицательный (б) эксцессы распределения.**

Точечные оценки параметров распределения. Пусть случайный эксперимент описывается случайной величиной *X*. Повторяя случайный эксперимент *n* раз, получаем последовательность наблюденных значений $\{x_1, x_2, \ldots, x_n\}$ – выборку объема *n* из генеральной совокупности случайной величины X.

Статистической оценкой параметра θ называется приближенное значение параметра, полученное на основе статистических (выборочных) данных $\tilde{\theta} = \tilde{\theta}(x_1, x_2, \ldots, x_n)$, где $\tilde{\theta}(x_1, x_2, \ldots, x_n)$ – выборочная характеристика, вычисляемая по результатам n наблюдений величины *X*, используемая в качестве оценки θ-характеристики генеральной совокупности. В качестве θ могут быть параметры $M(X), D(X)$, параметр распределения и т. д.

По статистическим данным нельзя получить точную оценку неизвестного параметра θ, но можно найти приближенную оценку. Более того, каждая выборка объема *n* из генеральной совокупности дает свою оценку одного и того же неизвестного параметра θ, т. е. для θ можно получить бесконечное множество его оценок. Поэтому оценку $\tilde{\theta} = \tilde{\theta}(x_1, x_2, \ldots, x_n)$ можно считать случайной величиной, а ее значение $\tilde{\theta}$, вычисленное по одной данной выборке, можно рассматривать как одно из ее возможных значений.

Различают точечные и интервальные оценки.

Точечная оценка параметра θ определяется одним числом $\tilde{\theta}$. Качество оценки $\tilde{\theta}$ устанавливается по приведенным ниже трем свойствам.

Несмещенность. Оценка $\tilde{\theta} = \tilde{\theta}(x_1, x_2, \ldots, x_n)$ неизвестного параметра θ генеральной совокупности называется *несмещенной*, если для фиксированного числа наблюдений *n* выполняется равенство $M(\tilde{\theta}) = \theta$.

Состоятельность. Оценка $\tilde{\theta} = \tilde{\theta}(x_1, x_2, \ldots, x_n)$, найденная по выборке объема n, называется *состоятельной*, если для любого $\varepsilon > 0$ выполняется равенство $\lim\limits_{n\to\infty} P\left(\left|\tilde{\theta} - \theta\right| < \varepsilon\right) = 1$, которое означает, что при увеличении объема *n* выборки значение $\tilde{\theta}$ сходится по вероятности к неизвестному параметру θ.

Эффективность. Несмещенная оценка $\tilde{\theta} = \tilde{\theta}(x_1, x_2, \ldots, x_n)$ неизвестного параметра θ называется *эффективной*, если среди всех подобных оценок того же параметра она имеет наименьшую дисперсию: $D(\tilde{\theta}) \to \min$.

Оценкой математического ожидания случайной величины *X* является ее выборочная средняя: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Она является несмещенной, состоятельной и эффективной.

Выборочная дисперсия $D_в = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ является *смещенной* оценкой для дисперсии случайной величины X.

Тест 7.13. Точечная оценка, математическое ожидание которой не равно оцениваемому параметру, называется:
1) смещенной;
2) несмещенной;
3) эффективной;
4) неэффективной;
5) состоятельной.

Тест 7.14. Выборочная средняя $\overline{X_в}$ является:
1) интервальной оценкой генеральной средней;
2) точечной оценкой генеральной дисперсии;
3) генеральной средней;
4) точечной оценкой генеральной средней;
5) интервальной оценкой генеральной дисперсии.

Тест 7.15. Выборочная дисперсия $D_в$ является:
1) интервальной оценкой генеральной средней;
2) точечной оценкой генеральной дисперсии;
3) генеральной средней;
4) точечной оценкой генеральной средней;
5) интервальной оценкой генеральной дисперсии.

Тест 7.16. Выборочное среднее квадратическое отклонение является:
1) интервальной оценкой генеральной средней;
2) точечной оценкой генеральной дисперсии;
3) генеральной средней;
4) точечной оценкой генеральной средней;
5) точечной оценкой генерального среднего квадратического отклонения.

Тест 7.17. Точечной оценкой не является:
1) выборочная средняя;
2) выборочная дисперсия;
3) выборочное среднее квадратическое отклонение;
4) доверительный интервал.

Величина $S^2 = \frac{n}{n-1}\sigma_в^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ является несмещенной, эффективной и состоятельной оценкой дисперсии случайной величины X и называется *несмещенной выборочной дисперсией*, а смещенная дисперсия – *генеральной дисперсией,* или *дисперсией генеральной совокупности*.

Величина $S = \sqrt{S^2}$ называется *стандартным отклонением*.

Пример 7.3. Выборочная дисперсия равна 2,7 при объеме выборки $n = 30$. Как оценить генеральную дисперсию?

Решение

Оценкой генеральной дисперсии является исправленная выборочная дисперсия $S^2 = \frac{30}{30-1}\cdot 2{,}7 \approx 2{,}7$.

Тест 7.18. Точечной оценкой генеральной дисперсии является:
1) исправленная выборочная дисперсия;
2) выборочная дисперсия;
3) выборочное среднее квадратическое отклонение;
4) доверительный интервал.

Тест 7.19. Выборочная дисперсия равна 2,5 при объеме выборки $n = 26$, тогда исправленная выборочная дисперсия равна:

1) $S^2 = \dfrac{26}{26-1} \cdot 2{,}5$;

2) $S^2 = \dfrac{26-1}{26} \cdot 2{,}5$;

3) $S^2 = \dfrac{2{,}5}{26}$.

Стандартная ошибка выборки определяется по формуле $\varepsilon = \dfrac{S}{\sqrt{n}}$.

7.5. Интервальные оценки неизвестных параметров

Более полный и надежный способ оценивания параметров распределений заключается в определении не единственного точечного значения, а интервала, который с заданной вероятностью содержит истинное значение оцениваемого параметра. Для заранее выбранного уровня значимости α, $0 < \alpha < 1$ по выборке определяются два числа $\tilde{\theta}_1$ и $\tilde{\theta}_2$, $\tilde{\theta}_1 < \tilde{\theta}_2$, между которыми с вероятностью $1-\alpha$ находится неизвестный параметр θ: $P(\tilde{\theta}_1 < \theta < \tilde{\theta}_2) = 1-\alpha$.

Число $p = 1-\alpha$ называется *доверительной вероятностью* (*надежностью*), $\tilde{\theta}_1$, $\tilde{\theta}_2$ – *доверительными* нижней и верхней *границами*. Величины $\tilde{\theta}_1, \tilde{\theta}_2$ определяются по результатам выборки, следовательно, являются случайными.

Если $\tilde{\theta}$ – точечная оценка неизвестного параметра θ, то $\tilde{\theta}_1 = \tilde{\theta} - \Delta$, $\theta_2 = \tilde{\theta} + \Delta$, где $\Delta (\Delta > 0)$ – *предельная ошибка* (*уровень надежности*) выборки, которая либо задается заранее, либо вычисляется.

На практике часто используются односторонние доверительные интервалы, которые определяются из условий $P(\theta < \tilde{\theta}_1) = 1-\alpha$ или $P(\theta < \tilde{\theta}_2) = 1-\alpha$ и называются *правосторонними* и *левосторонними* соответственно.

Длина доверительного интервала, характеризующая точность интервальной оценки, зависит от объема выборки n и надежности p. При увеличении n длина доверительного интервала уменьшается, а с приближением надежности к 1 – увеличивается. В качестве $p = 1-\alpha$ принимают значения 0,9; 0,95; 0,99, что соответствует 90, 95, 99%-ым доверительным интервалам соответственно.

Задача определения доверительного интервала может быть решена только тогда, когда удается найти закон распределения случайной величины, используемой в качестве оценки. В общем случае этот закон зависит от самого неизвестного параметра. Однако иногда удается перейти от оценки $\tilde{\theta}$ к таким функциям выборочных значений, закон распределения которых зависит только от объема выборки n и закона распределения случайной величины и не зависит от неизве-стных параметров.

Пусть выборка $\{x_1, x_2, \ldots, x_n\}$ произведена из генеральной совокупности значений нормально распределенной с параметрами $M(X) = a$ и $D(X) = \sigma^2$ случайной величины X, т.е. $X \sim N(a; \sigma^2)$.

Доверительный интервал для математического ожидания $M(X)$ *при известной дисперсии* $D(X) = \sigma^2$ с доверительной вероятностью $p = 1-\alpha$ имеет вид:

$$\left(\bar{x} - t\frac{\sigma}{\sqrt{n}}, \bar{x} + t\frac{\sigma}{\sqrt{n}}\right) \text{ или } (\bar{x} - \delta, \bar{x} + \delta).$$

Здесь $t = t\left(\dfrac{1-\alpha}{2}\right)$ – квантиль порядка $\dfrac{1-\alpha}{2}$ нормального распределения, заданного функцией распределения вероятностей

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy, \text{ т. е. } \Phi(x) = \frac{1-\alpha}{2}$$

Выражение $\delta = t\dfrac{\sigma}{\sqrt{n}}$ – точность оценки. Из соотношения $\delta = t\dfrac{\sigma}{\sqrt{n}}$ находится минимальный объем $n = \left(\dfrac{t \cdot \sigma}{\delta}\right)^2$ выборки, который обеспечивает заданную точность δ.

Число σ, как правило, неизвестно, поэтому его заменяют приближенным значением: $\sigma \approx S$.

Пример 7.4. Автомат наполняет пакеты с чипсами. Установлено по выборочным данным, что стандартное отклонение веса пакетов $\sigma_в = 10$ г. Среднее значение веса пакета $\bar{x}_в = 249$ г (при n = 36). В каком интервале с надежностью 95% лежит истинное значение веса, считая что вес подчиняется нормальному распределению $X \sim N(m\,;10)$.

Решение

Для определения 95% доверительного интервала найдем критическую точку $t_{\frac{1-\alpha}{2}} = t_{0,475}$ из уравнения

$$\Phi(t) = \frac{0{,}95}{2} = 0{,}475.$$

По таблицам функции Лапласа $t_{0,475} = 1{,}96$. Тогда доверительный интервал имеет вид $\left(249 - 1{,}96 \cdot \frac{10}{6};\, 249 + 1{,}96 \cdot \frac{10}{6}\right)$.

Тест 7.20. Доверительный интервал для оценки математического ожидания при выборочной средней $\overline{X_в} = 14$ и точности оценки $\delta = 1{,}5$ имеет вид:

1) (12,5; 15,5);
2) (14,5; 17);
3) (0;14);
4) (14; 28).

Вопросы для самоконтроля

1. Какие числовые характеристики выборки относятся к показателям положения?
2. Какие числовые характеристики выборки относятся к показателям разброса?
3. Какой показатель характеризует симметрию распределения?
4. Что называется выборочным средним, выборочной модой, выборочной медианой, выборочной дисперсией?
5. Каким условиям должны удовлетворять точечные оценки?
6. Какие оценки называются интервальными?
7. Что называется доверительной вероятностью?

Ответы на тестовые задания

Номер теста	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	7.10
Правильный ответ	2	1	1	2	4	1	1	2	3	2

Номер теста	7.11	7.12	7.13	7.14	7.15	7.16	7.17	7.18	7.19	7.20
Правильный ответ	4	3	1	4	2	5	4	1	1	1

8. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ О ПАРАМЕТРАХ НОРМАЛЬНО РАСПРЕДЕЛЕННОЙ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ

Пусть X – наблюдаемая дискретная или непрерывная случайная величина. *Статистической гипотезой* H называется утверждение, в котором высказывается предположение относительно параметров или вида закона распределения случайной величины X. Гипотеза H называется *простой*, если она содержит только одно предположение, а гипотеза, которая состоит из конечного или бесконечного числа простых гипотез, называется *сложной*. Гипотезы о значениях параметров известного распределения случайной величины X называются *параметрическими*. *Непараметрическими* называются гипотезы, сформулированные относительно вида закона распределения случайной величины X. Основная выдвинутая гипотеза называется *нулевой* H_0.

Гипотеза, противоречащая нулевой гипотезе H_0, называется *альтернативной (конкурирующей) гипотезой* H_1. Выбор альтернативной гипотезы определяется конкретной формулировкой задачи.

Правило, по которому принимается решение принять или отклонить гипотезу H_0, называется *критерием K*. Случайная величина K, с помощью которой принимается решение о принятии или отклонении нулевой гипотезы, называется с*татистикой K* критерия K. Проверка статистической гипотезы основывается на *принципе отношения правдоподобия*: маловероятные события считаются невозможными, а события, имеющие большую вероятность, считаются достоверными. Зафиксируем некоторую малую вероятность α – *уровень значимости*. Пусть W – множество значений статистики K, $W_K \subseteq W$ – такое подмножество, для которого $P(K \in W_K | H_0 \text{ верна}) = \alpha$.

Наблюдаемым значением $K_{набл}$ называется значение статистики K, вычисленное по выборке $\{x_1; x_2; ...; x_n\}$ из генеральной совокупности случайной величины X. Правило проверки статистических гипотез состоит в следующем:

• если $K_{набл} \in W_K$, то гипотеза H_0 отклоняется с вероятностью $P(K_{набл} \in W_K | H_0 \text{ верна}) = \alpha$,

• если $K_{набл} \in W \setminus W_K$, то гипотеза H_0 принимается с вероятностью $P(K_{набл} \in W \setminus W_K | H_0 \text{ верна}) = 1 - \alpha$.

Критерий, основанный на использовании заранее заданного уровня значимости α, называется *критерием значимости*. Вероятность $p = 1 - \alpha$ называется *доверительной вероятностью*.

Критической областью W_K называется совокупность значений статистики K, при которых нулевую гипотезу отвергают. *Областью принятия* гипотезы называется совокупность значений W/W_K статистики K, при которых нулевая гипотеза принимается. *Критическими точками* K_{kp} называются точки, отделяющие критическую область W_K от области принятия гипотезы W/W_K. Уровень значимости α определяет «размер» критической области W_K. Положение критической области на множестве статистики W зависит от вида нулевой и альтернативной гипотез. Возможны три вида расположения критической области :

- правосторонняя критическая область $(K_{кр\ пр}; +\infty)$;
- левосторонняя критическая область $(-\infty; K_{кр\ лев})$;
- двусторонняя критическая область $(-\infty; K_{кр\ лев}) \cup (K_{кр\ пр}; +\infty)$.

Точки $K_{кр\ пр}$, $K_{кр\ лев}$ определяются в зависимости от вида закона распределения статистики K при выбранном уровне значимости α.

Выбор между гипотезами H_0 и H_1 может сопровождаться ошибками двух родов. *Ошибка первого рода* состоит в том, что будет отвергнута правильная нулевая гипотеза H_0. Вероятность ошибки первого рода равна уровню значимости α:

$$\alpha = P(\text{отвергнуть } H_0 | \, H_0 \text{ верна}).$$

Ошибка второго рода состоит в том, что будет принята неправильная нулевая гипотеза H_0 с вероятностью β:

$$\beta = П(\text{принять } H_0 | \, H_0 \text{ неверна }).$$

Вероятности ошибок должны быть малыми и выбираться заранее.

При проверке гипотезы возникает одна из следующих четырех ситуаций, приведенных в табл. 5.

Таблица 5. **Ошибки первого и второго рода**

Результаты проверки гипотезы	*Возможные состояния гипотезы*	
	H_0 верна	*H_0 неверна*
Гипотеза H_0 отклоняется	Ошибка первого рода	Правильный вывод
Гипотеза H_0 принимается	Правильный вывод	Ошибка второго рода

Мощностью критерия называется вероятность попадания статистики K в критическую область W_K при условии, что справедлива конкурирующая (альтернативная) гипотеза H_1. Мощность критерия равна вероятности γ правильного отклонения нулевой гипотезы H_0:

$$\gamma = 1 - \beta = P(\text{отвергнуть} H_0 | H_1 \text{ верна}).$$

Поскольку критическая область W_K определяется по-разному на заданном уровне значимости α, то она выбирается так, чтобы мощность критерия γ была возможно большей:

$$P(K_{набл} \in W \setminus W_K | H_1 \text{ верна}) \to \max.$$

Чем больше мощность критерия, тем меньше вероятность принятия неверной гипотезы. Мощность критерия позволяет выбрать оптимальную статистику K для проверки гипотезы среди возможных статистик критерия.

На практике в качестве статистики K чаще всего используются специально подобранные случайные величины, распределения которых известны:

- Z (стандартизированное нормальное распределение);
- t (распределение Стьюдента);

- χ^2 (закон Пирсона χ^2);
- F (распределение Фишера).

Высказываемые в ходе решения задач гипотезы условно подразделяются на следующие типы:

- о виде закона распределения исследуемой случайной величины;
- об однородности двух или нескольких выборок;
- о числовых значениях параметров исследуемого признака;
- об общем виде зависимости, существующей между компонентами исследуемого многомерного признака.

Общая схема проверки статистических гипотез. Несмотря на разнообразие гипотез и применяемых статистик, проверка статистических гипотез может быть проведена в виде следующей общей схемы:

1. На основании выборочных данных выдвигаются нулевая гипотеза H_0 и альтернативная ей гипотеза H_1.

2. Выбирается уровень значимости α (в практических задачах пользуются стандартными значениями уровня значимости: $\alpha = 0{,}1$; 0,05; 0,025; 0,01; 0,005; 0,001).

3. Выбирается статистика K, имеющая известный закон распределения.

4. Вычисляется наблюдаемое значение статистики $K_{набл}$ по выборочным данным.

5. Определяется вид критической области из условия $P(K \in W_K \mid H_0) = \alpha$ и область принятия гипнозы в зависимости от формулировки альтернативной гипотезы.

6. Принимается статистическое решение: если $K_{набл}$ попадает в критическую область, то нулевая гипотеза H_0 отвергается, в противном случае H_0 принимается.

Тест 8.1. Критическая область для проверки гипотезы H_0 имеет вид $(K_{кр};+\infty)$. Гипотеза будет отвергнута, если:

1) $K_{набл} < K_{кр}$;
2) $K_{набл} > K_{кр}$;
3) $K_{набл} = K_{кр}$;
4) $K_{набл} = 0$;
5) $K_{кр} = 0$.

Тест 8.2. Критическая область для проверки гипотезы H_0 имеет вид: $(-\infty; K_{кр})$. Гипотеза будет отвергнута, если:

1) $K_{набл} < K_{кр}$;
2) $K_{набл} > K_{кр}$;
3) $K_{набл} = K_{кр}$;
4) $K_{набл} = 0$;
5) $K_{кр} = 0$

Тест 8.3. Область принятия гипотезы H_0 имеет вид $(-K_{кр}; K_{кр})$. Гипотеза H_0 будет принята, если:

1) $K_{набл} < K_{кр}$;
2) $K_{набл} > K_{кр}$;
3) $K_{набл} = K_{кр}$;
4) $-K_{кр} < K_{набл} < K_{кр}$;
5) $K_{кр} = 0$

Изучение реальных процессов предполагает получение не только прогнозной оценки исследуемого показателя, но и количественной характеристики степени влияния на него различных факторов, а также оценки возможных последствий их изменений в будущем. В результате опыта проводятся наблюдения над целым рядом случайных величин. При этом возникает задача изучения взаимосвязи между случайными величинами, которая решается в три этапа:

- проводится оценка существенности влияния одного фактора на другой с помощью дисперсионного анализа;
- проводится численная оценка связи с помощью корреляционного анализа;

* строятся функциональные зависимости посредством регрессионного анализа.

Дисперсионный анализ служит для статистического установления влияния отдельных факторов на изменчивость какого-либо признака, значения которого могут быть получены опытным путем в виде выборки из генеральной совокупности случайной величины X. Под *факторами* понимаются различные независимые показатели, количество которых может быть различным. Конкретная реализация фактора A называется *уровнем* (*группой*) этого фактора. В зависимости от количества факторов различают однофакторный и многофакторный дисперсионный анализ. Величина X называется *результативным признаком* (*фактором*) Y. Идея дисперсионного анализа состоит в том, что дисперсия признака Y разлагается на сумму дисперсии, вызванной влиянием факторов, дисперсии, вызванной взаимодействием факторов и случайной дисперсии, вызванной неучтенными случайными факторами. Затем указанные дисперсии сравниваются и проверяются по статистическим критериям.

Однофакторный дисперсионный анализ позволяет статистически обосновать степень влияния на результативный признак Y одного фактора А.

Дисперсионный анализ рассматривает влияние двух независимых факторов А и В на изменчивость результативного признака Y.

Тест 8.4. Пусть в результате измерения величины M получено значение X, и пусть на процесс измерения влияют случайные независимые факторы A и B. Тогда для оценки значимости факторов A и B применяют:

1) однофакторный дисперсионный анализ;
2) двухфакторный дисперсионный анализ;
3) корреляционный анализ;
4) регрессионный анализ;
5) трехфакторный дисперсионный анализ.

Вопросы для самоконтроля

1. Что называется критерием и мощностью критерия?
2. Что определяет уровень значимости гипотезы?
3. Что такое критическая область критерия?
4. Как найти доверительную вероятность статистического критерия?
5. Какие виды ошибок могут быть при проверке гипотез?
6. Что называется мощностью критерия?
7. Какие статистики используются при проверке гипотез о законе распределения?
8. В чем заключается суть дисперсионного анализа?
9. В каком случае используется однофакторный дисперсионный анализ?
10. В чем разница между однофакторным и двухфакторным анализами?

Ответы на тестовые задания

Номер теста	8.1	8.2	8.3	8.4
Правильный ответ	2	1	4	2

9. ИССЛЕДОВАНИЕ ВЗАИМОСВЯЗИ МЕЖДУ ПРИЗНАКАМИ

9.1. Ковариация и корреляция

В естественных науках важной задачей является анализ зависимостей между изучаемыми величинами. При обработке и использовании статистических данных с целью получения как научных, так и практических выводов вызывает интерес, как изменяется один признак при изменении другого. Эти зависимости могут быть функциональными, стохастическими. В функциональных зависимостях каждому значению одной переменной соответствует единственное значение другой, т. е. результативный признак полностью определяется факторным. *Стохастической* называется зависимость, при которой изменение одной из величин влечет изменение распределения другой. Факторный признак в таких связях не полностью определяет результативный признак, а влияет только на среднее значение, при этом отдельные результаты могут противоречить установленной связи.

В статистике изучаются наблюденные данные случайных величин, поэтому стохастическая зависимость называется *статистической,* или *корреляционной.*

Задачами корреляционного анализа являются:

- измерение степени связи;
- отбор факторов, оказывающих наибольшее влияние на результативный признак на основании степени связности между признаками;
- обнаружение неизвестных причинных связей.

Для оценки тесноты и вида связи между случайными величинами используются показатели ковариации и корреляции.

Пусть $(x_1;y_1)$, $(x_2;y_2)$, …, $(x_n;y_n)$ – выборка из генеральной совокупности двумерной случайной величины $(X;Y)$, описывающей случайный эксперимент.

Выборочной ковариацией $\operatorname{cov}(X;Y)$ называется среднее произведений отклонений значений выборок $\{x_1;x_2;...;x_n\}$ и $\{y_1;y_2;...;y_n\}$ от их средних $\bar{x}$, $\bar{y}$: $\operatorname{cov}(X;Y)=\dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\overline{xy}-\bar{x}\,\bar{y}$.

Ковариация характеризует рассеивание значений выборок X и Y, а также линейную связь между ними.

Выборочной ковариационной матрицей называется матрица вида

$$\begin{pmatrix}\operatorname{cov}(X;X) & \operatorname{cov}(X;Y)\\ \operatorname{cov}(Y;X) & \operatorname{cov}(Y;Y)\end{pmatrix}.$$

Очевидно, что $\operatorname{cov}(X;X)=\sigma_X^2$, $\operatorname{cov}(Y;Y)=\sigma_Y^2$, где для оценки σ_x^2, σ_y^2 используют исправленные выборочные дисперсии S_x^2, S_y^2.

Выборочная ковариационная матрица устанавливает взаимосвязь между выборками X и Y из генеральных совокупностей:

- при $\operatorname{cov}(X,Y)>0$ большим значениям выборки X соответствуют большие значения выборки Y;
- при $\operatorname{cov}(X,Y)<0$ большим значениям выборки X соответствуют меньшие значения выборки Y;
- при $\operatorname{cov}(X,Y)\to 0$ данные выборок X и Y не связаны.

Выборочным коэффициентом корреляции r_{XY} случайных величин X и Y, между которыми предполагается линейная корреляционная связь, называется величина, определяемая по формуле

$$r_{XY}=\frac{\operatorname{cov}(X;Y)}{\sigma_X\sigma_Y}=\frac{\overline{xy}-\bar{x}\,\bar{y}}{S_xS_y}.$$

Качественная оценка тесноты связи между величинами выявляется по шкале Чеддока (табл. 6)

Таблица 6. **Шкала Чеддока**

Теснота связи	Значение коэффициента корреляции при наличии	
	прямой связи	обратной связи
Слабая	0,1–0,3	(–0,1)–(–0,3)
Умеренная	0,3–0,5	(–0,3)–(–0,5)
Заметная	0,5–0,7	(–0,5)–(–0,7)
Высокая	0,7–0,9	(–0,7)–(–0,9)
Весьма высокая	0,9–0,99	(–0,9)–(–0,99)

Выборочной корреляционной матрицей называется матрица вида $\begin{pmatrix}1 & r_{XY}\\ r_{YX} & 1\end{pmatrix}$.

Выборочная корреляционная матрица также устанавливает взаимосвязь наборов выборочных данных по величине:

1) при $0<r_{XY}<1$ большим значениям выборки X соответствуют большие значения выборки Y;

2) при $-1<r_{XY}<0$ большим значениям выборки X соответствуют меньшие значения выборки Y (или наоборот);

3) при $r_{XY}=0$ данные двух диапазонов некоррелированы;

4) при $|r_{XY}|=1$ существует линейная функциональная зависимость между выборочными значениями X и Y.

При исследовании связи между несколькими случайными величии-нами находятся выборочные коэффициенты ковариации и корреляции между парами всех исследуемых величин и строятся соответ-ствующие ковариационные и корреляционные матрицы.

Тест 9.1. В формуле для вычисления коэффициента линейной корреляции $r = \dfrac{\overline{xy} - ?\,\overline{y}}{?\cdot\overline{S_y}}$ вместо «?» надо поставить:

1) $\overline{x}\,; \overline{S}_x$
2) $\overline{y}\,; \overline{S}_y$
3) $\overline{x}\,, \left(\sigma_b^2\right)_x$;
4) $1, \left(\sigma_b\right)_y^2$;
5) $0,\ \dfrac{\overline{y}}{\overline{x}}$.

Тест 9.2. Если признаки X, Y независимы, то коэффициент корреляции r_{xy} равен:

1) $r_{xy} = 0$;
2) $r_{xy} = -1$;
3) $r_{xy} = 1$;
4) $r_{xy} < \dfrac{1}{2}$;
5) $r_{xy} > -\dfrac{1}{2}$.

Тест 9.3. Коэффициент корреляции $r_{xy} = -1$, тогда связь между признаками:

1) прямая;
2) обратная линейная;
3) XY независимы;
4) прямая линейная.

Тест 9.4. Если признаки X и Y линейно зависимы, причем наблюдается прямая зависимость, то:

1) $r_{xy} = 0$;
2) $r_{xy} = -1$;
3) $r_{xy} = 1$;
4) $r_{xy} < \dfrac{1}{2}$;
5) $r_{xy} > -\dfrac{1}{2}$.

9.2. Регрессия

Наряду с корреляционным анализом проводится *регрессионный анализ*, который заключается в определении формы связи зависимой случайной величины y с независимыми случайными величинами $X_1, X_2, \ldots, X_n$.

Форма связи результативного признака y с факторами $X_1, X_2, \ldots, X_n$ называется *уравнением регрессии*. В зависимости от типа выбранного уравнения различают *линейную* и *нелинейную регрессию* (квадратичную, экспоненциальную, логарифмическую и т. д.).

В зависимости от числа взаимосвязанных признаков различают *парную* и *множественную регрессии*. Если исследуется связь между двумя признаками (результативным и факторным), то регрессия называется *парной*, если между тремя и более признаками – *множественной* (*многофакторной*) *регрессией*.

Тест 9.5. Уравнение $y = 5 + 0{,}2x_1 + 1{,}7x_2$ определяет следующий вид регрессии:

1) множественный линейный;
2) множественный нелинейный;
3) парный линейный;
4) парный нелинейный.

При изучении регрессии следует придерживаться определенной последовательности этапов.

Этап 1. Установление формы зависимости. Пусть в результате наблюдений двумерной случайной величины $(X;Y)$ получены данные, представляющие собой совокупность точек $(x_1;y_1)$, $(x_2;y_2)$, …, $(x_n;y_n)$. Графическое изображение этих точек в плоскости Oxy представляет собой корреляционное поле (диаграмму рассеяния). Диаграмма рассеяния позволяет произвести визуальный анализ эмпирических данных и графически определить вид функции регрессии.

Тест 9.6. Результаты измерений признаков X и Y изображены в виде точек $(x_i; y_j)$, на корреляционном поле в виде рис. 20.

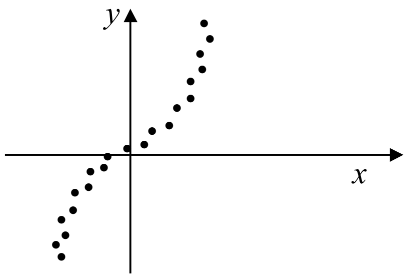

Рис. 20

Тогда связь между признаками является:
1) линейной;
2) квадратической;
3) экспоненциальной;
4) логарифмической;
5) кубической.

Тест 9.7. Результаты измерений признаков X и Y изображены в виде точек $(x_i; y_j)$, на корреляционном поле в виде рис. 20. Зависимость между признаками определяется уравнением:
1) $y = ax + b$;
2) $y = ax^2 + bx + c$;
3) $y = ae^{bx}$;
4) $y = a\ln x$;
5) $y = ax^3 + bx^2 + cx + d$.

Этап 2. Определение параметров (коэффициентов) уравнения регрессии.

Параметры уравнения регрессии определяются с помощью метода наименьших квадратов.

Этап 3. Проверка общего качества уравнения регрессии. Для определения величины степени стохастической взаимосвязи результативного признака и факторов необходимо знать следующие дисперсии:

- общую дисперсию результативного признака y, отображающую влияние как основных, так и остаточных факторов: $\sigma_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$, где $\bar{y}$ – выборочное среднее значение результативного признака y по выборке $\{y_1, y_2, ..., y_n\}$;
- факторную дисперсию результативного признака y, отображающую влияние только основных факторов: $\sigma_\Phi^2 = \frac{1}{n}\sum_{i=1}^{n}(\tilde{y}_i - \bar{y})^2$, где $\tilde{y}_i = f(x_i)$значения, найденные по уравнению регрессии;
- остаточную дисперсию результативного признака, отображающую влияние только остаточных факторов:

$$\sigma_O^2 = \frac{1}{n-(m+1)}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2 .$$

При корреляционной связи результативного признака и факторов выполняется соотношение: $\sigma_\Phi^2 < \sigma_y^2$, при этом $\sigma_y^2 = \sigma_\Phi^2 + \sigma_O^2$.

Для анализа общего качества уравнения линейной регрессии обычно используется *множественный коэффициент детерминации* R^2, называемый также *квадратом коэффициента множественной корреляции.* Множественный коэффициент детерминации рассчитывается по формуле $R^2 = \frac{\sigma_\Phi^2}{\sigma_y^2}$ и определяет долю разброса результативного признака, обусловленную изменением факторных признаков, входящих в многофакторную модель. Чем теснее линейная связь между признаками, тем ближе коэффициент детерминации к единице. Однако, при достаточно близком к единице коэффициенте детерминации не всегда наблюдается тесная взаимосвязь между случайными величинами. Поэтому необходимы дополнительные исследования.

Этап 4. Проверка статистической значимости каждого коэффициента уравнения регрессии и определение их доверительных интервалов.

Проверяются соответствующие гипотезы

Вопросы для самоконтроля

1. Что называется гипотезой?
2. Какая гипотеза называется нулевой, альтернативной, простой и сложной?
3. Какая зависимость называется стохастической?
4. Каковы основные задачи корреляционного анализа?
5. Какая зависимость называется корреляционной?
6. Какая величина называется выборочной ковариацией? Что она характеризует?
7. Как по коэффициенту корреляции оценить тесноту связи между случайными величинами?
8. В чем разница между корреляционным и регрессионным анализом?
9. Как построить корреляционное поле?
10. Какое уравнение называется уравнением регрессии?
11. Какой вид имеет уравнение линейной регрессии?
12. Как связаны общая и факторная дисперсия при корреляционной зависимости результативного признака и факторов?
13. Что используется для анализа общего качества уравнения линейной регрессии?

Ответы на тестовые задания

Номер теста	9.1	9.2	9.3	9.4	9.5	9.6
Правильный ответ	1	1	2	3	1	5

СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

Белько, И. В. Теория вероятностей и математическая статистика. Примеры и задачи : учеб. пособие для вузов / И. В. Белько, Г. П. Свирид ; под ред. К. К. Кузьмина. – Минск : Новое знание, 2002. – 250 с.

Булдык, Г. М. Теория вероятностей и математическая статистика : учеб. пособие для экон. специальностей вузов / Г. М. Булдык. – Минск : Выш. шк., 1989. – 284 с.

Гмурман, В. Е. Теория вероятностей и математическая статистика : учеб. пособие для вузов / В. Е. Гмурман. – М. : Высш. шк., 2003. – 479 с.

Гмурман, В. Е. Руководство к решению задач по теории вероятностей и математической статистике : учеб. пособие для вузов / В. Е. Гурман. – М. : Высш. шк., 2003. – 405 с.

Гусак, А. А. Задачи и упражнения по высшей математике : учеб. пособие для вузов. В 2 кн. Кн. 2. – Минск : Выш. шк., 1988. – 228 с.

Жевняк, Р. М. Теория вероятностей и математическая статистика : учеб. пособие для вузов / Р. М. Жевняк, А. А. Карпук, В. Т. Унукович. – Минск : Харвест, 2000. – 384 с.

Колемаев, В. А. Теория вероятностей и математическая статистика : учеб. для вузов / В. А. Колемаев, В. Н. Калинина ; под ред. В. А. Колемаева. – М. : ИНФРА-М, 2001. – 302 с.

Фигурин, В. А. Теория вероятностей и математическая статистика : учеб. пособие для вузов / Ф. А. Фигурин, В. В. Оболонкин. – Минск : Новое знание, 2000. – 208 с.

СОДЕРЖАНИЕ

Учебное издание

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Пособие по подготовке к тестированию для студентов заочной формы обучения экономических специальностей

Авторы-составители:
Авдашкова Людмила Павловна,
Калмыкова Тамара Федоровна,
Кузменкова Инна Анатольевна и др.

Редактор Е. Г. Привалова
Технический редактор Н. Н. Короедова
Компьютерная верстка И. А. Козлова

Подписано в печать 25.09.09. Бумага типографская № 1.
Формат 60 × 84 1/16. Гарнитура Таймс. Ризография.
Усл. печ. л. 7,90. Уч.-изд. л. 8,00. Тираж 500 экз.
Заказ №

Учреждение образования
«Белорусский торгово-экономический университет
потребительской кооперации».
246029, г. Гомель, просп. Октября, 50.

ЛИ № 02330/0494302 от 04.03.2009 г.

Отпечатано в учреждении образования
«Белорусский торгово-экономический университет
потребительской кооперации».
246029, г. Гомель, просп. Октября, 50.